Invariance and Body-Order Compose Additively: Minimax Rates on $\mathrm{SO}(3)^n$

Zheshuo Li **Zhengxiong Li**
Department of Computer Science and Engineering
University of Colorado Denver
Denver, CO 80217, USA
{zheshuo.li,zhengxiong.li}@ucdenver.edu

Abstract

Two structural priors—global invariance under a group action and low body-order—are widely used in equivariant architectures on product Lie groups such as $\mathrm{SO}(3)^n$. Whether their statistical gains are fundamental, changing the exponent of the minimax rate, or merely constant-factor, changing only the leading coefficient, has remained open. We show that these gains are statistically fundamental: both priors reduce the effective dimension governing the minimax rate, not merely the leading constant. Specifically, we establish matching upper and lower minimax rates for four function classes on $\mathrm{SO}(3)^n$, one for each combination of the two priors. The matching lower bound relies on a pure-S pullback construction and a combinatorial packing across the $\binom{n}{K}$ top-order subsets. The resulting effective-dimension formula decomposes additively: invariance contributes a fixed reduction of 3, while body-order truncation at order K contributes a reduction of $3(n-K)$ that grows with system size. An adaptive sample-splitting procedure selects the body-order from data, achieving the best in-family risk up to an $O(\sqrt{\log N/N})$ remainder with no N-dependent tuning. Five experiments with spectral estimators corroborate the theory's predicted structural and operational consequences, supporting the four-class hierarchy on synthetic targets and yielding 2–17$\times$ sample-efficiency gains on a Maier–Saupe-type benchmark from liquid-crystal theory.

1 Introduction

Functions of multiple orientations on $\mathrm{SO}(3)^n$ arise naturally in robotics, rigid-body mechanics, and molecular science. In molecular science, they describe liquid-crystal order parameters [23, 10] and protein–protein interaction energies [21]; the same product structure appears in multi-body articulated systems in robotics and rigid-body mechanics. Two structural priors recur across these domains: global invariance—the function depends only on relative orientations, not on a global reference frame—and low body-order—the function decomposes into pairwise, triplet, or at most K-body interactions. Equivariant network architectures and body-ordered expansions exploit these priors to great empirical effect [31, 3, 12]. Yet their combined statistical value on product Lie groups—the interaction of invariance and body-order at the rate level—remains unresolved.

The distinction matters. If structural priors only improve constants, their statistical benefit would eventually saturate as datasets grow. If they reduce the governing dimension, data requirements change at the exponent level, and the priors become increasingly valuable as the problem grows. More formally, we ask: *For nonparametric estimation on $\mathrm{SO}(3)^n$, do structural priors such as global invariance and body-order truncation merely improve finite-sample constants, or do they instead reduce the effective statistical dimension that governs the minimax rate?*

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

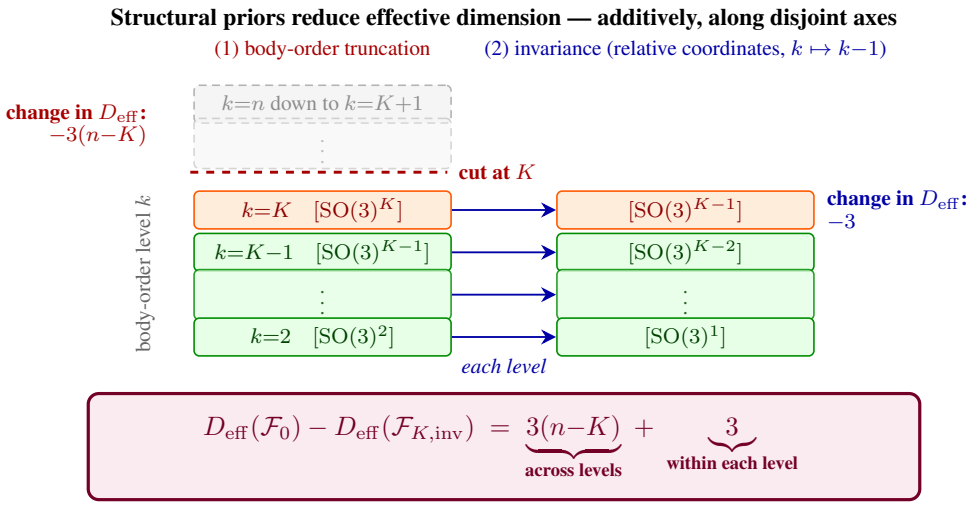

Figure 1: **Effective dimension decomposes additively: 3 from invariance, $3(n-K)$ from body-order truncation.** Body-order truncation removes whole levels $k > K$ (horizontal cut, left); invariance takes each surviving level from $\mathrm{SO}(3)^k$ to $\mathrm{SO}(3)^{k-1}$ by passing to relative coordinates (blue arrows, right). The rate-governing top order ($k{=}K$) contributes -3 to D_{eff} under invariance; truncation lowers the top order from n to K, reducing its dimension by $3(n-K)$.

As Figure 1 shows, these priors admit an **exact additive effective-dimension accounting**: invariance contributes a size-independent reduction of 3 in effective dimension, body-order truncation at order K contributes an n-dependent reduction of $3(n-K)$, and the reductions compose additively because they act on disjoint axes of complexity.

Beyond the rate formula, the theory yields concrete, testable consequences. On a physically motivated Maier–Saupe benchmark, incorporating both priors saves 2–17$\times$ samples relative to weaker structural classes (Table 2). An adaptive sample-splitting procedure selects the body-order from data with an oracle-type guarantee, requiring no N-dependent tuning.

Our contributions are:

- **First minimax rates for combined structural priors on product Lie groups.** To our knowledge, we give the first matching upper and lower bounds for four function classes on $\mathrm{SO}(3)^n$ (Theorem 1), showing that structural priors reduce the effective statistical dimension governing the minimax rate, not merely the constant. Its lower bound rests on two ingredients absent from prior analyses of single-quotient invariance [30] and Euclidean analysis of variance (ANOVA) [5]: *a pure-S pullback construction* (Theorem 18) and *a combinatorial packing* across the $\binom{n}{K}$ top-order subsets (Appendix G).

- **Additive structural decomposition.** Theorem 1's matching bounds reveal that the invariance-induced reduction (3) and the body-order reduction ($3(n-K)$) in effective dimension compose additively, along disjoint axes of complexity (Corollary 2)—a structural consequence absent from prior single-quotient invariance or Euclidean ANOVA analyses.

- **Adaptive extension.** A sample-splitting procedure selects the body-order K from data, achieving the best in-family risk plus an $O(\sqrt{\log N/N})$ remainder (Proposition 3), under the same smoothness condition required by Theorem 1.

- **Empirical corroboration and operational consequences.** Five experiments probe the theory at three levels—a per-class quantitative exponent check, structural consequences (hierarchy and n-dependence), and operational consequences (adaptive selection and 2–17$\times$ sample-efficiency gains on a physical Maier–Saupe benchmark)—each matching its claim type.

2 Related Work

Nonparametric estimation on compact groups and manifolds. Classical minimax theory characterizes optimal estimation rates on $\mathbb{R}^d$ [28, 32], with extensions to spheres [18] and general Riemannian manifolds [6]. Harmonic-analytic and group-representation machinery for probability and statistics on compact groups is well established [11, 1]. The closest analogue on the domain side is Jeon et al. [17], who analyze nonparametric regression on compact Lie groups with measurement errors but do not consider invariance, body-order truncation, or product-group structure. Minimax rates on product Lie groups $\mathbf{G} = G^n$ specifically have received little systematic study.

Equivariant deep learning. Group equivariant convolutional networks [9] introduced a systematic framework for building symmetry-respecting architectures, extended to continuous groups via steerable CNNs [36], Lie group convolutions [15], and the general theory of equivariant neural networks on compact groups [20, 24, 8]. Equivariant networks for 3D geometry [31, 16, 27] and body-ordered architectures such as ACE [12], MACE [3], and NequIP [4] achieve strong empirical performance across molecular property prediction and materials modeling.

Statistical value of invariance. Recent work has begun to quantify the sample-complexity gains from invariance: Elesedy and Zaidi [14] proved a strict generalization benefit for equivariant linear models, Elesedy [13] extended this to kernel methods, and Mei et al. [25] and Bietti et al. [7] analyzed invariant kernel regression. The closest analogue on the invariance side is Tahmasebi and Jegelka [30], who established the exact sample-complexity gain from invariance for kernel ridge regression on compact manifolds under the action of a compact Lie group, reducing the effective dimension by that of the group orbit (up to a factor related to the volume of the quotient space). Applied to $\mathrm{SO}(3)^n$ with diagonal $\mathrm{SO}(3)$ action, their framework yields a reduction of 3 in effective dimension—matching our $\mathcal{F}_{\mathrm{inv}}$ rate (Proposition 32)—but treats invariance as a single global quotient and does not accommodate an ANOVA decomposition across body-order levels. In contrast, once an ANOVA decomposition is layered on top, the matching lower bound requires a pure-S pullback (Theorem 18) and a combinatorial packing across the $\binom{n}{K}$ top-order subsets (Appendix G), yielding additive rather than interactive composition.

Body-order expansion and ANOVA decomposition. The ANOVA decomposition on $\mathbb{R}^d$ [35, 22, 29] and sparse additive models [26] both leverage low interaction order for statistical efficiency. The atomic cluster expansion (ACE) in materials science [12, 2] provides a body-ordered representation of interatomic potentials. The closest analogue on the body-order side is Bhattacharya et al. [5], who establish rates for ANOVA-type function classes on Euclidean domains with diverging dimension. Extending this analysis to $\mathrm{SO}(3)^n$ requires three ingredients: representation-theoretic mode counting via Peter–Weyl; quotienting the coupled diagonal $\mathrm{SO}(3)$ action jointly rather than factor by factor; and constructing pure-S ANOVA components compatible with that action for the matching lower bound.

Taken together, these lines of work leave open a common question: on product Lie groups such as $\mathrm{SO}(3)^n$, can we precisely quantify the minimax-rate benefit of combining symmetry with body-order structure? The present work addresses this gap (Theorem 1).

3 Theory

3.1 Setup

Observation model and minimax risk. We observe N i.i.d. pairs (X_i, Y_i) where $X_i = (g_{i,1}, \dots, g_{i,n}) \sim \mu^{\otimes n}$ is drawn uniformly from $\mathbf{G} = \mathrm{SO}(3)^n$, with μ the normalized Haar measure on $\mathrm{SO}(3)$, and $Y_i = f^*(X_i) + \xi_i$ with $\xi_i \sim \mathcal{N}(0, \sigma^2)$ independent of X_i. Error is measured in squared $L^2(\mu^{\otimes n})$ norm, and the minimax risk of a class $\mathcal{F}$ is

$$\mathcal{R}_N(\mathcal{F}) \;:=\; \inf_{\widehat{f}} \sup_{f^* \in \mathcal{F}} \mathbb{E}_{f^*} \big\| \widehat{f} - f^* \big\|^2_{L^2(\mu^{\otimes n})},$$

where the infimum ranges over all measurable estimators built from the N observations and the expectation is over design and noise. The worst case is over targets, while the estimator is chosen to minimize that worst-case risk. A rate r_N is *minimax-optimal* for $\mathcal{F}$ if $c\, r_N \le \mathcal{R}_N(\mathcal{F}) \le C\, r_N$ with constants depending only on n, K, s, R, σ.

Table 1: The four function classes, their structural priors, the effective domain, the effective statistical dimension D_{eff}, and the minimax rate (Theorem 1). This table is the paper's central result: each structural prior changes the exponent, not just the constant.

Class	Structure	Effective domain	D_{eff}	Minimax rate
$\mathcal{F}_0$	none	$\mathrm{SO}(3)^n$	$3n$	$N^{-2s/(2s+3n)}$
$\mathcal{F}_{\text{inv}}$	invariance only	$\mathrm{SO}(3)^{n-1}$ (relative coords)	$3(n-1)$	$N^{-2s/(2s+3(n-1))}$
$\mathcal{F}_K$	body-order $\leq K$ only	$\mathrm{SO}(3)^K$ per K-subset	$3K$	$\binom{n}{K}/N\big)^{2s/(2s+3K)}$
$\mathcal{F}_{K,\text{inv}}$	both	$\mathrm{SO}(3)^{K-1}$ per K-subset	$3(K-1)$	$\binom{n}{K}/N\big)^{2s/(2s+3(K-1))}$

Harmonic analysis and smoothness. By the Peter–Weyl theorem [19, 11], $L^2(\mathbf{G})$ decomposes into irreducible representations indexed by $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n) \in \mathbb{Z}^n_{\geq 0}$, with an orthonormal basis given by normalized products of Wigner-D matrix entries [33]; write $\widehat{f}(\boldsymbol{\ell})$ for the matrix of Fourier coefficients of f at $\boldsymbol{\ell}$, $d_{\boldsymbol{\ell}} = \prod_i (2\ell_i + 1)$ for the dimension of the representation, and $\lambda_{\boldsymbol{\ell}} = \sum_i \ell_i(\ell_i + 1)$ for its Casimir eigenvalue. The Sobolev norm of order $s > 0$ is

$$\|f\|^2_{H^s(\mathbf{G})} \;:=\; \sum_{\boldsymbol{\ell}} (1 + \lambda_{\boldsymbol{\ell}})^s \, d_{\boldsymbol{\ell}} \left\|\widehat{f}(\boldsymbol{\ell})\right\|^2_{\text{HS}},$$

so larger s forces the Fourier coefficients to decay faster, i.e. a smoother target. The target class is the Sobolev ball $\mathcal{W}^s(\mathbf{G}) = \{f : \|f\|_{H^s} \leq R\}$; the same definition applies on $\mathrm{SO}(3)^m$ for any m. The number of basis functions with $\lambda_{\boldsymbol{\ell}} \leq \Lambda$ on $\mathrm{SO}(3)^m$ grows as $\Theta(\Lambda^{3m/2})$ (Weyl counting; Lemma 5), which is where the factor $3m$ in the effective dimension comes from.

Four function classes. We define four Sobolev balls, one for each combination of two structural priors:

- **Global invariance**: $f(hg_1, \ldots, hg_n) = f(g_1, \ldots, g_n)$ for all $h \in \mathrm{SO}(3)$. The function depends only on relative orientations $g_1^{-1} g_j$.
- **Body-order truncation at order** K: $f = \sum_{|S| \leq K} f_S$ where f_S depends only on $\{g_i : i \in S\}$ and the components satisfy an ANOVA orthogonality condition.

The four classes represent frame-independent interactions of order at most K ($\mathcal{F}_{K,\text{inv}}$), interactions of order at most K in absolute coordinates, as under an external field ($\mathcal{F}_K$), frame-independent dependence with unrestricted body order ($\mathcal{F}_{\text{inv}}$), and unrestricted dependence ($\mathcal{F}_0$). They satisfy $\mathcal{F}_{K,\text{inv}} \subseteq \mathcal{F}_K \cap \mathcal{F}_{\text{inv}} \subseteq \mathcal{F}_0$; the two single-prior classes are incomparable when $K < n$. Invariance removes 3 from D_{eff} on either branch, while the relative difficulty of the single-prior classes depends on K versus $n-1$ (with equal effective dimensions at $K = n-1$). For n=6, K=2, the effective dimensions are $3, 6, 15, 18$ for $\mathcal{F}_{K,\text{inv}}, \mathcal{F}_K, \mathcal{F}_{\text{inv}}, \mathcal{F}_0$, respectively. For fixed n, K, the $\binom{n}{K}$ top-order subsets govern the body-ordered minimax rates, affecting their scale but not the N-exponent. Under global invariance, pure one-body ANOVA components vanish, so K=2 is the first nonconstant interaction order.

Relative coordinates and the pure-S pullback. For $S = \{i_1, \ldots, i_k\}$ with $|S| = k \geq 2$, an invariant k-body component can be written as $f_S(\mathbf{g}) = \varphi_S(g_{i_1}^{-1} g_{i_2}, \ldots, g_{i_1}^{-1} g_{i_k})$ for a *descended representative* φ_S on $\mathrm{SO}(3)^{k-1}$, with comparable Sobolev norms (Appendix D); the relative domain has dimension $3(k-1)$. Let $\mathcal{V}_S$ denote the invariant ANOVA subspace with Fourier support exactly S (Appendix B). Pullback to $\mathbf{G}$ may have a nonzero projection onto $\bigoplus_{S' \subsetneq S} \mathcal{V}_{S'}$, which we call *lower-order leakage*. For full-support inputs, the pure-S pullback removes this leakage by applying $P^{\perp}_{\Delta}$ (defined below) before pullback (Theorem 18); on such inputs the projection is the identity at k=2 and can be nontrivial at $k \geq 3$ (Appendix N).

Estimator for $\mathcal{F}_{K,\text{inv}}$. The matching upper bound for $\mathcal{F}_{K,\text{inv}}$ is attained by an empirical Fourier projection in relative coordinates. For a k-subset $S = \{i_1, \ldots, i_k\}$ let $q_S : \mathbf{G} \to \mathrm{SO}(3)^{k-1}$, $q_S(\mathbf{g}) = z_S := (g_{i_1}^{-1} g_{i_2}, \ldots, g_{i_1}^{-1} g_{i_k})$, be the relative-coordinate map, which carries $\mu^{\otimes n}$ to product Haar measure, and define the pullback $T_S = q_S^*$ by $(T_S\varphi)(\mathbf{g}) = \varphi(q_S(\mathbf{g}))$, so that $f_S = T_S \varphi_S$ above. Let A_Δ average a function on $\mathrm{SO}(3)^{k-1}$ over simultaneous left rotations, $(A_\Delta \varphi)(r_2, \ldots, r_k) = \int \varphi(hr_2, \ldots, hr_k)\, d\mu(h)$, and set $P^{\perp}_{\Delta} := I - A_\Delta$. For each $k \in \{2, \ldots, K\}$ and each k-subset S:

(i) form the relative coordinates $z_S^{(\alpha)} = q_S(X_\alpha)$ of each sample; (ii) take a real L^2-orthonormal basis $\{\phi_{S,j}\}_{j=1}^{M_S}$ of the modes on $\mathrm{SO}(3)^{k-1}$ with Casimir eigenvalue at most Λ_k that have full support (every relative coordinate active) and lie in $\mathrm{range}(P_\Delta^\perp)$; for large Λ_k the *mode count* M_S, the number of retained basis functions, satisfies $M_S \asymp \Lambda_k^{3(k-1)/2}$ (Lemma 19); (iii) estimate each coefficient by an empirical average, $\widehat{a}_{S,j} = N^{-1}\sum_{\alpha=1}^N Y_\alpha\, \phi_{S,j}(z_S^{(\alpha)})$; (iv) pull back and assemble,

$$\widehat{f} \;=\; \bar{Y} + \sum_{k=2}^{K} \sum_{|S|=k} \sum_{j=1}^{M_S} \widehat{a}_{S,j}\, T_S\phi_{S,j}, \qquad \Lambda_k \asymp \big(N/\tbinom{n}{k}\big)^{2/(2s+3(k-1))},$$

with $\bar{Y} = N^{-1}\sum_\alpha Y_\alpha$ (Definition 23). The level-wise cutoff Λ_k balances bias against variance, and the full-support and $P_\Delta^\perp$ constraints keep each component estimator inside the pure S-body subspace (Section 3.2). The other three classes use the same recipe on their own domains: subset-by-subset on $\mathrm{SO}(3)^k$ in absolute coordinates for $\mathcal{F}_K$, and a single expansion on $\mathrm{SO}(3)^{n-1}$ or $\mathrm{SO}(3)^n$ for $\mathcal{F}_{\mathrm{inv}}$ and $\mathcal{F}_0$ (Appendix H).

3.2 Minimax rates for the four classes

Theorem 1 (Minimax rates)**.** *Under the class-specific smoothness condition $s > D_{\mathrm{eff}}/2$—i.e., $s > 3(K-1)/2$ for $\mathcal{F}_{K,\mathrm{inv}}$, $s > 3K/2$ for $\mathcal{F}_K$, $s > 3(n-1)/2$ for $\mathcal{F}_{\mathrm{inv}}$, and $s > 3n/2$ for $\mathcal{F}_0$—all four minimax rates in Table 1 are tight: matching upper and lower bounds hold up to constants depending only on n, K, s, R, σ.*

As in classical Sobolev embedding, the condition $s > D_{\mathrm{eff}}/2$ ensures that the target is uniformly bounded, which the variance analysis needs (Lemma 6; Lemma 22 for $\mathcal{F}_{K,\mathrm{inv}}$).

The key message is comparative: moving from $\mathcal{F}_0$ to $\mathcal{F}_{K,\mathrm{inv}}$ changes the rate exponent from $2s/(2s+3n)$ to $2s/(2s+3(K-1))$—a fundamental improvement that grows with n. We prove the $\mathcal{F}_{K,\mathrm{inv}}$ case in full (upper bound in Appendix F, lower bound in Appendix G) and obtain the remaining three rates by parallel arguments specialized to their reduced domains (Appendix H).

Proof sketch. **Upper bound.** For each k-body subset S, the component estimator has bias $\lesssim (1+\Lambda_k)^{-s}\|\varphi_S\|_{H^s}^2$ and variance $\lesssim \Lambda_k^{3(k-1)/2}/N$, where φ_S is the pure relative-coordinate representative of f_S (Corollary 21). The level-wise cutoff specified above depends on N, n, k, s and the fixed model parameters, but not on the unknown target. Sobolev norm equivalence and the global budget $\sum_{k=2}^K \sum_{|S|=k}\|\varphi_S\|_{H^s}^2 \le R^2$ (Appendix D) allow component risks to be summed across subsets; the top body-order $k{=}K$ then dominates, yielding the stated rate.

Lower bound. To prove the matching lower bound, we construct $m \asymp \Lambda^{3(K-1)/2}$ orthonormal functions in each pure-S invariant ANOVA subspace $\mathcal{V}_S$, $|S| = K$. A full-support relative-coordinate function φ need not satisfy $T_S\varphi \in \mathcal{V}_S$: its pullback may contain lower-order components (Lemma 17(i)). For such φ, the lower-order projection of $T_S\varphi$ is determined by $A_\Delta\varphi$ (Lemma 17(ii)), so $T_S(P_\Delta^\perp\varphi) \in \mathcal{V}_S$ (Theorem 18). The removed diagonal-invariant subspace has dimension $O(\Lambda^{3(K-2)/2})$, leaving $\asymp \Lambda^{3(K-1)/2}$ pure modes (Lemma 19). These modes yield mutually orthogonal families across the $\binom{n}{K}$ top-order subsets; Varshamov–Gilbert packing and Fano's inequality give the matching lower bound (Appendix G).

3.3 Additive dimension decomposition

A key structural consequence of Theorem 1 is that the dimension reductions from the two priors compose exactly:

Corollary 2 (Additive decomposition)**.** *The total dimension reduction from $\mathcal{F}_0$ to $\mathcal{F}_{K,\mathrm{inv}}$ decomposes additively:*

$$D_{\mathrm{eff}}(\mathcal{F}_0) - D_{\mathrm{eff}}(\mathcal{F}_{K,\mathrm{inv}}) \;=\; 3n - 3(K-1) \;=\; \underbrace{3}_{\textit{invariance}} \;+\; \underbrace{3(n-K)}_{\textit{body-order}}. \tag{1}$$

Theorem 1 establishes the minimax-sharp rates for all four classes; Corollary 2 characterizes the combined effect of the two priors. The upper bound for $\mathcal{F}_{K,\mathrm{inv}}$ alone only implies $D_{\mathrm{eff}}(\mathcal{F}_{K,\mathrm{inv}}) \le$

$3(K-1)$, leaving open the possibility of an additional reduction from their interaction. The matching lower bound rules out such a reduction, establishing the exact additive decomposition $3 + 3(n-K)$ with no interaction term. This additivity concerns effective dimensions, not the nonlinear rate exponents $2s/(2s + D_{\text{eff}})$.

3.4 Adaptive body-order selection

When K^* is unknown, we select among a pre-specified finite family of candidates: for each body order $k \in \{2, \dots, K_{\max}\}$ a degree cutoff $\ell_{\max,k}$ is fixed independently of N, and candidate K' retains the per-factor-truncated pure modes of all orders $k \le K'$. For $K' \ge 2$ the retained relative-coordinate modes have Casimir eigenvalues at most $\Lambda_{K'} = \max_{k \le K'}(k-1)\,\ell_{\max,k}(\ell_{\max,k}+1)$, and fitting the candidate requires $O\big(N \sum_{k=2}^{K'} \binom{n}{k} (1 + \ell_{\max,k})^{3(k-1)}\big)$ basis evaluations. We give an oracle inequality for selection within this fixed family, with no smoothness assumption beyond Theorem 1; Appendix J.5 treats an N-dependent alternative under the stronger condition $s > 3(K_{\max}-1)/2$.

Proposition 3 (Adaptive oracle inequality over a fixed-cutoff family)**.** *Fix $K_{\max} \ge 2$ and degree cutoffs $\ell_{\max,k}$, $2 \le k \le K_{\max}$, independent of N; let $M_{K'}$ be the resulting finite mode count of candidate K'. Let $\widehat{f}_N$ be the sample-splitting estimator of Definition 38. Then for all $f^* \in \mathcal{F}_{K^*,\text{inv}}$ with $2 \le K^* \le K_{\max}$:*

$$\mathbb{E}\big[\|\widehat{f}_N - f^*\|_{L^2}^2\big] \;\le\; \min_{K' \le K_{\max}} R_N(K') \;+\; \tau_N, \qquad \tau_N \;\le\; C\sqrt{\frac{\log(2K_{\max}N)}{N}}, \tag{2}$$

where $R_N(K') := \mathbb{E}\|\widehat{f}^{(K')} - f^*\|_{L^2}^2$ is the risk of the K'-candidate, $\min_{K'} R_N(K')$ is the in-family oracle over the pre-specified candidates, and C depends only on $n, K_{\max}, s, R, \sigma$ and the cutoffs $\{\ell_{\max,k}\}$. If the true-order candidate satisfies $R_N(K^*) \le C_0\, r_N(K^*)$ for a constant C_0 independent of N, where $r_N(K^*) = \big(\binom{n}{K^*}/N\big)^{2s/(2s+3(K^*-1))}$, then $\mathbb{E}\big[\|\widehat{f}_N - f^*\|_{L^2}^2\big] \le C_0\, r_N(K^*) + \tau_N$.

Sample splitting gives the in-family oracle coefficient one, with an additive selection cost bounded by $O(\sqrt{\log N/N})$. This bound does not establish minimax-rate adaptation: a fixed dictionary may retain approximation bias, and even under the condition above, the available bound on τ_N is asymptotically larger than $r_N(K^*)$ when $s > 3(K^*-1)/2$.

4 Experiments

The five main experiments use body-order $K{=}2$, the most common low-body-order regime in practice, with smoothness $s{=}2$, noise variance $\sigma^2{=}0.1$, and uniform Haar design on $\mathrm{SO}(3)^n$; Theorem 1 itself holds for general K and s. At $s{=}2$ only $\mathcal{F}_{K,\text{inv}}$ satisfies the smoothness condition of Theorem 1; the other three classes are reported as pre-asymptotic checks (Section 5). Appendix N checks the pure-S projection at $K{=}3$, the first body order where it becomes nontrivial. The four estimators use finite Wigner-D spectral dictionaries with Sobolev-weighted coefficient shrinkage (Appendix K), allowing us to compare the structural priors without architectural or optimization effects; protocol details and robustness checks are in Appendices K, L, M, and N.

Roadmap. Experiment 4.1 compares empirical and theoretical learning-curve slopes. Experiment 4.2 examines how these slopes change with n, guided by the dimension prediction of Corollary 2. Experiment 4.3 compares the four estimators on shared targets. Experiment 4.4 evaluates body-order selection, while Experiment 4.5 measures sample-efficiency gains on the Maier–Saupe benchmark, followed by a neural bridge.

4.1 Empirical rate exponents are consistent with the predicted dimension formula

For each class, we fit its estimator on a class-matched target with Sobolev-decaying coefficients ($n{=}5$, $K{=}2$; Appendix K) at sample sizes from $N{=}500$ to 50,000. We compare the empirical log-log slope $\hat{\alpha}$ of risk against N with the theoretical slope $\alpha_{\text{theory}} = -2s/(2s + D_{\text{eff}})$. Both slopes are negative; Figure 2(b) shows their absolute values.

As shown in Figure 2, three of four classes—$\mathcal{F}_{K,\text{inv}}$ ($D_{\text{eff}}{=}3$, theory -0.571), $\mathcal{F}_K$ ($D_{\text{eff}}{=}6$, -0.400), and $\mathcal{F}_{\text{inv}}$ ($D_{\text{eff}}{=}12$, -0.250)—yield empirical slopes within 0.05 of theory. $\mathcal{F}_0$ ($D_{\text{eff}}{=}15$, theory -0.211) shows a steeper slope of -0.38. For the $\mathcal{F}_0$ estimator, we impose a fixed per-factor degree

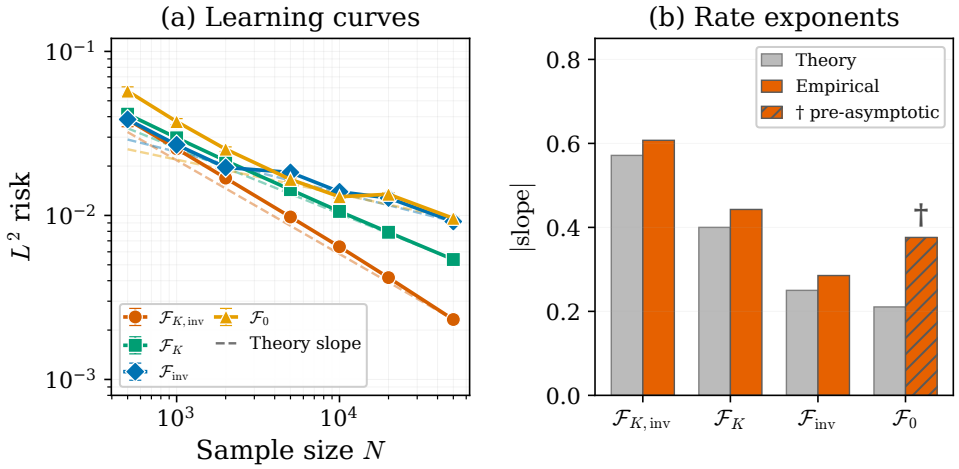

Figure 2: Learning curves (panel a) and empirical vs. theoretical rate exponents (panel b) for the four classes, each with its own class-matched ground truth and spectral estimator; panel (b) plots slope magnitudes $|\alpha|$. Three of the four empirical slope magnitudes are within 0.05 of the theoretical values $2s/(2s + D_{\mathrm{eff}})$; $\mathcal{F}_0$ (D_{eff}=15) remains pre-asymptotic over the tested range. Error bars in panel (a) show ± 1 standard deviation across 30 repetitions; the same convention is used in all learning-curve panels throughout the paper.

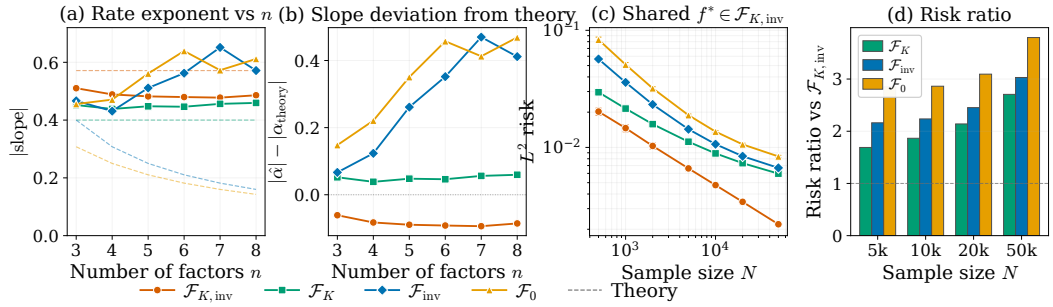

Figure 3: Empirical structural comparisons (Experiments 4.2–4.3). *Panels (a,b): Experiment 4.2.* (a) Empirical and theoretical rate exponents as n grows; (b) the resulting slope-magnitude gaps $|\hat{\alpha}| - |\alpha_{\mathrm{theory}}|$. Fixed-$D_{\mathrm{eff}}$ classes ($\mathcal{F}_{K,\mathrm{inv}}$, $\mathcal{F}_K$) show n-independent gaps near zero; growing-D_{eff} classes ($\mathcal{F}_{\mathrm{inv}}$, $\mathcal{F}_0$) show gaps that widen with n. *Panels (c,d): Experiment 4.3.* Across three targets in $\mathcal{F}_{K,\mathrm{inv}}$ (n=6), each shared across estimators within a trial: (c) the ordering $\mathcal{F}_{K,\mathrm{inv}} < \mathcal{F}_K < \mathcal{F}_{\mathrm{inv}} < \mathcal{F}_0$ holds at every sample size; (d) the risk ratios $R(\mathcal{F}_c)/R(\mathcal{F}_{K,\mathrm{inv}})$ increase with N over the displayed range.

cap $\ell_{\max}$=1 for computational tractability, in addition to the joint Casimir cutoff (Appendix K); under these finite cutoffs, $\mathcal{F}_0$ has not reached its bias-variance balance over the tested range, so the observed -0.38 is a pre-asymptotic effect, as marked in panel (b).

4.2 n-dependence separates fixed-D from growing-D classes

$\mathcal{F}_{K,\mathrm{inv}}$ and $\mathcal{F}_K$ have fixed D_{eff} $(3(K-1), 3K)$; $\mathcal{F}_{\mathrm{inv}}$ and $\mathcal{F}_0$ have n-dependent D_{eff} $(3(n-1), 3n)$. We sweep n from 3 to 8 and plot the slope-magnitude gap $|\hat{\alpha}| - |\alpha_{\mathrm{theory}}|$ per class (positive = empirical slope steeper than minimax).

As shown in Figure 3(b), $\mathcal{F}_{K,\mathrm{inv}}$ and $\mathcal{F}_K$ keep n-independent offsets near zero (-0.09 and $+0.05$): their D_{eff} (3 and 6) does not depend on n. For $\mathcal{F}_{\mathrm{inv}}$ and $\mathcal{F}_0$ the gap widens with n: their theoretical slope magnitudes fall from 0.40 and 0.31 at n=3 to 0.16 and 0.14 at n=8 as $D_{\mathrm{eff}} = 3(n-1)$ and $3n$ grow (Figure 3(a)), while the empirical slopes under fixed truncation stay comparatively stable.

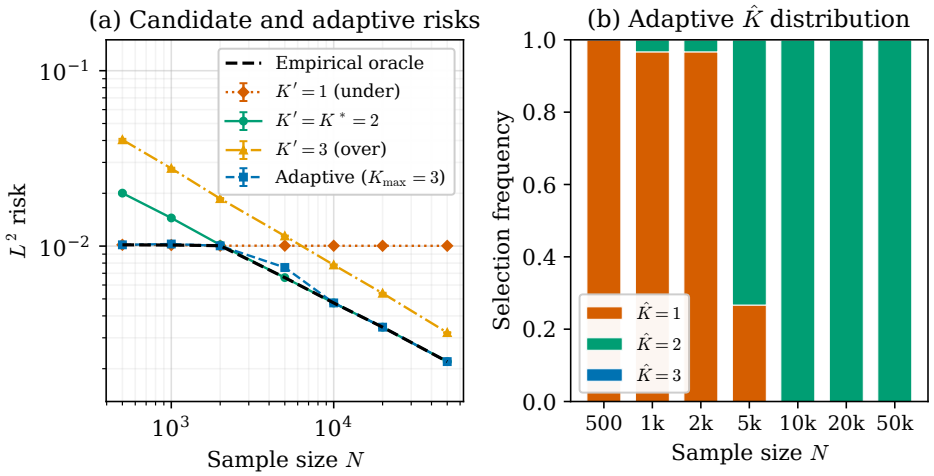

Figure 4: Panel (a): raw L^2 risks for three candidates—K'=1 (Under-specified, orange dotted), K'=K^*=2 (green), K'=3 (Over-specified, amber dash-dot)—together with the Adaptive estimator ($K_{\max}$=3, blue dashed) and the empirical oracle (black dashed), the pointwise minimum over the three candidate risks. Panel (b): distribution of selected body-order $\hat{K}$ across trials. At small N, the Under-specified candidate achieves lower risk than the K'=K^*=2 candidate (variance advantage outweighs bias) and the Adaptive curve tracks it; at large N, the K'=K^*=2 candidate becomes preferable and the Adaptive curve merges with it.

4.3 Four-estimator risk ordering on shared targets

Experiments 4.1–4.2 use class-matched targets. Here we generate three targets in $\mathcal{F}_{K,\text{inv}}$ (n=6) and evaluate all four estimators on the same noisy observations within each trial. Every class contains these targets, but the finite dictionaries of the three less structured classes omit some of their higher-degree modes (Appendix K), as in Experiment 4.5.

As shown in Figure 3(c), the observed ordering $\mathcal{F}_{K,\text{inv}} < \mathcal{F}_K < \mathcal{F}_{\text{inv}} < \mathcal{F}_0$ holds for the mean risks at every tested sample size (N=500 to 50,000), and panel (d) shows the risk ratios $R(\mathcal{F}_c)/R(\mathcal{F}_{K,\text{inv}})$ increasing over N=5,000 to 50,000, from 1.69 to 2.71 for $\mathcal{F}_K$, 2.16 to 3.03 for $\mathcal{F}_{\text{inv}}$, and 2.85 to 3.79 for $\mathcal{F}_0$. Part of this growing gap also comes from modes missing in the $\mathcal{F}_K$ dictionary, so the risk ratios alone do not measure the minimax exponent gap.

4.4 Adaptive body-order selection achieves near-oracle risk

We evaluate the stabilized implementation (Appendix K) of the sample-splitting scheme of Definition 38 with n=6, K^*=2, $K_{\max}$=3, and degree cutoffs $\ell_{\max,2}$=5 for pair modes and $\ell_{\max,3}$=1 for triplet modes, held fixed across all N. The data is split evenly: $N/2$ samples fit the candidates at $K' \in \{1, 2, 3\}$ and $N/2$ score them; the retained coefficients undergo the Sobolev-weighted shrinkage used throughout Section 4, and the selected candidate is refit on all N samples. We track the ratio of the adaptive risk to the empirical oracle, the pointwise minimum over the three candidates' mean test risks with each candidate fit on all N samples, and the selected order $\hat{K}$ as a secondary diagnostic.

The adaptive-to-oracle risk ratio, shown in Figure 4a as the gap between the Adaptive (blue) and empirical oracle (black) curves, peaks at 1.14 near the transition $N \approx 5{,}000$ and remains close to 1 for $N \geq 10{,}000$, indicating small empirical selection overhead within this fixed candidate family. As a secondary diagnostic (Figure 4b), the selected order $\hat{K}$ transitions from 1 to 2 as N grows past $\sim$5,000, and $\hat{K}$=3 is never chosen, indicating that the holdout criterion avoids unnecessary dimensionality without requiring knowledge of K^*.

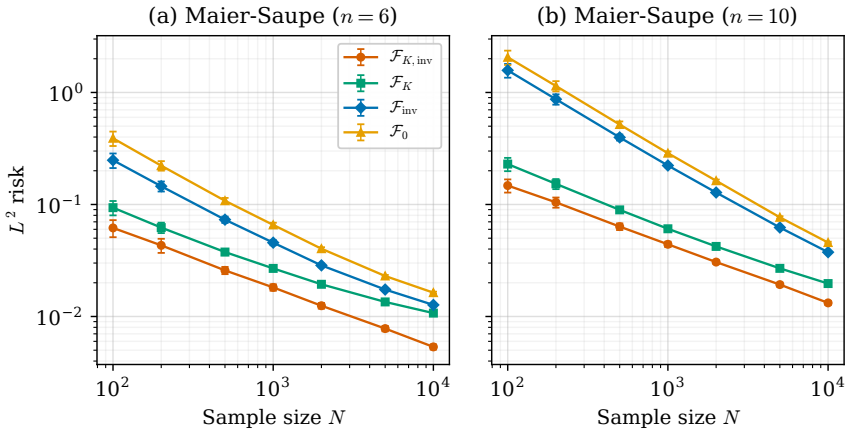

Figure 5: Learning curves on the Maier–Saupe benchmark at n=6 (left) and n=10 (right); both panels use the same logarithmic L^2-risk scale. The four-class hierarchy $\mathcal{F}_{K,\mathrm{inv}} < \mathcal{F}_K < \mathcal{F}_{\mathrm{inv}} < \mathcal{F}_0$ holds at both n, with the gap between $\mathcal{F}_{K,\mathrm{inv}}$ and the weaker classes widening at n=10—consistent with the n-dependent body-order reduction.

Table 2: Sample-efficiency multipliers on the Maier–Saupe benchmark. Each entry: samples needed by class C to match $\mathcal{F}_{K,\mathrm{inv}}$ at anchor N, divided by anchor N ("—": not reached).

	$n = 6$			$n = 10$		
Anchor N	$\mathcal{F}_K$	$\mathcal{F}_{\mathrm{inv}}$	$\mathcal{F}_0$	$\mathcal{F}_K$	$\mathcal{F}_{\mathrm{inv}}$	$\mathcal{F}_0$
200	1.9×	5.4×	9.1×	1.9×	12.9×	17.2×
500	2.2×	4.8×	8.3×	1.8×	9.8×	12.9×
1000	2.4×	4.6×	8.0×	1.8×	8.0×	—

4.5 Structural priors save 2–17× samples on a physical benchmark

This experiment reports operational consequences of the dimension hierarchy—sample-efficiency multipliers on a physically motivated benchmark—rather than a re-estimation of the minimax exponent. Prior experiments use synthetic H^s targets; here we test whether the effective-dimension hierarchy persists on a physically motivated target outside our construction pipeline. We use a Maier–Saupe-type pair potential on SO(3), $f^*(\mathbf{g}) = -c_n \sum_{i<j} \big[P_2(x_{ij}) + 0.3\, P_4(x_{ij})\big]$, where P_2, P_4 are Legendre polynomials, $x_{ij} = \frac{1}{2}\big(\mathrm{tr}(g_i^{-1} g_j) - 1\big)$ is the cosine of the relative rotation angle (rather than the director angle of classical Maier–Saupe theory), and $c_n > 0$ normalizes the target standard deviation to approximately 0.1. This globally invariant pairwise target lies in $\mathcal{F}_{K,\mathrm{inv}}$ with K=2 and has per-pair Peter–Weyl support in degrees $\ell \le 4$. The four estimators use class-specific finite cutoffs (Appendix K); the dictionaries of the less structured classes omit some higher-degree target modes (matching the $\mathcal{F}_{K,\mathrm{inv}}$ resolution in absolute coordinates would square the per-pair mode count), so the observed risk gaps can reflect both structural mode-count differences and cutoff-induced approximation error. We test at n=6 and n=10 to probe whether the body-order advantage widens with system size, as predicted by Corollary 2.

As shown in Figure 5, the hierarchy holds at both n; Table 2 translates it into an operational quantity—how many more samples each weaker prior needs to match $\mathcal{F}_{K,\mathrm{inv}}$ at a given budget. Two patterns stand out. First, the $\mathcal{F}_K/\mathcal{F}_{K,\mathrm{inv}}$ ratio stays near 2× regardless of n, consistent with the size-independent invariance reduction (the fixed 3-term in Corollary 2). Second, the $\mathcal{F}_{\mathrm{inv}}$ and $\mathcal{F}_0$ ratios grow from n=6 to n=10 (at N=500: 4.8× → 9.8× for $\mathcal{F}_{\mathrm{inv}}$, 8.3× → 12.9× for $\mathcal{F}_0$), consistent with the n-dependent body-order reduction (the $3(n-K)$-term). The multipliers are nearly unchanged under mild model misspecification (λ=0.3 three-body contamination), and the ordering is preserved at higher smoothness (s=3), where the multipliers shrink; see Appendices L and M.

Neural bridge. Appendix O trains four parameter-matched MLP learners that encode none, one, or both priors on the same Maier–Saupe benchmark. The effective dimensions order the four classes as

$3 < 6 < 15 < 18$ at n=6 and $3 < 6 < 27 < 30$ at n=10, and the mean bridge curves show the same strict ordering from N=5,000 and N=2,000 respectively, with the $\mathcal{F}_{K,\mathrm{inv}}$-like learner lowest at every N. These learners are not the estimators Theorem 1 analyzes; their fitted slopes are descriptive, not rate exponents. Appendix P reports an architectural control for the both-prior learner.

5 Discussion and Conclusion

Main takeaway. Our analysis resolves the dichotomy posed in the introduction: the two structural priors that pervade equivariant machine learning—global invariance and body-order truncation—do not merely improve finite-sample constants on $\mathrm{SO}(3)^n$. They reduce the rate-governing effective dimension by 3 (invariance) and $3(n-K)$ (body-order truncation); these reductions add exactly and improve the minimax convergence exponent. Practically, invariance is worth incorporating whenever the target is reference-frame independent, body-order truncation grows in value with n, and the two should be combined whenever both apply. We emphasize that the additive decomposition itself is a theorem-level result (Corollary 2); our experiments corroborate its predicted consequences—rate exponents, four-class hierarchy, n-dependence, and sample-efficiency multipliers—rather than the additivity statement directly.

Broader significance. For equivariant deep learning, our results provide one statistical mechanism consistent with the sample efficiency observed in body-ordered equivariant architectures such as MACE and NequIP, rather than attributing it to a finite-sample coincidence. These bounds make this structural benefit precise, and whether practical neural architectures attain these rates under optimization remains an open question. For statistical learning theory, this is the first minimax characterization of combined structural priors on non-Euclidean product domains, generalizing the ANOVA + Sobolev rate literature from $\mathbb{R}^n$ to compact Lie group products. The lower-bound argument combines quotient geometry, pure-mode counting, and information-theoretic packing, suggesting possible extensions to other compact group products with compatible body-order decompositions.

Limitations. We highlight four limitations. **Estimator class:** the lower bounds of Theorem 1 apply to all measurable estimators, including kernel ridge regression (KRR). We establish the matching upper bounds using spectral projection; Remark 31 outlines a route toward matching KRR guarantees. Whether practical equivariant neural networks attain these rates under optimization remains an open question. **Body-order and smoothness:** Experiments use K=2 and s=2 (with s=3 appendix robustness). At s=2, the theorem's condition $s > D_{\mathrm{eff}}/2$ holds only for $\mathcal{F}_{K,\mathrm{inv}}$, so the three larger classes operate outside the strict asymptotic regime and are reported as pre-asymptotic checks; we use s=2 deliberately to expose the hierarchy and n-dependence within feasible sample sizes. The K=3 study of Appendix N is a construction check, not a rate validation; quantitative rate validation at higher body orders remains future work. **Empirical breadth:** experiments include one physically motivated benchmark (Maier–Saupe); broader empirical coverage on additional non-synthetic orientation-dependent targets would strengthen the practical scope of the study. **Domain specificity:** our theory is established on $\mathrm{SO}(3)^n$. For $\mathrm{SO}(m)^n$, the same geometric argument suggests candidate effective dimensions $d_m n$, $d_m(n-1)$, $d_m K$, and $d_m(K-1)$, where $d_m = \dim \mathrm{SO}(m) = m(m-1)/2$, and hence an additive reduction of $d_m + d_m(n-K)$. The quotient-space identity used in our mode-count argument extends to compact Lie groups, but establishing the corresponding minimax rates would require pure-S constructions, mode counting and tensor-product multiplicities for highest-weight representations, and Sobolev–Casimir bounds under pullback. Proving this extension, as well as extensions to $\mathrm{SE}(3)^n$ and to non-uniform body-order truncation, remains future work.

Conclusion. We provide the first exact minimax characterization of combined structural priors on $\mathrm{SO}(3)^n$: global invariance and body-order truncation reduce the effective statistical dimension additively, by 3 and $3(n-K)$ respectively, with no interaction term. The result establishes, at the minimax-rate level, the statistical value of the priors that underlie equivariant architectures, and suggests extensions to other compact Lie groups, while leaving noncompact settings such as $\mathrm{SE}(3)^n$ and rate attainment by practical equivariant architectures as open questions.

Acknowledgments and Disclosure of Funding

This work is partially supported by NSF Awards #2426470, #2436428, and Colorado OEDIT AI Grant Award. The authors declare no competing interests.

References

[1] David Applebaum. *Probability on compact Lie groups*, volume 70. Springer, 2014.

[2] Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Phys. Rev. B*, 87:184115, May 2013.

[3] Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: higher order equivariant message passing neural networks for fast and accurate force fields. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

[4] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, May 2022.

[5] Sohom Bhattacharya, Jianqing Fan, and Debarghya Mukherjee. Deep neural networks for nonparametric interaction models with diverging dimension. *The Annals of Statistics*, 52(6):2738–2766, 2024.

[6] Peter J. Bickel and Bo Li. *Local Polynomial Regression on Unknown Manifolds*, pages 177–186. Institute of Mathematical Statistics, 2007.

[7] Alberto Bietti, Luca Venturi, and Joan Bruna. On the sample complexity of learning under geometric stability. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 18673–18684, 2021.

[8] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Velickovic. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *CoRR*, abs/2104.13478, 2021.

[9] Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, JMLR Workshop and Conference Proceedings, pages 2990–2999. JMLR.org, 2016.

[10] Pierre-Gilles de Gennes and Jacques Prost. *The Physics of Liquid Crystals*. Number 83 in International Series of Monographs on Physics. Oxford University Press, Oxford, 2nd edition, 1993.

[11] Persi Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *IMS Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, 1988.

[12] Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B*, 99:014104, Jan 2019.

[13] Bryn Elesedy. Provably strict generalisation benefit for invariance in kernel methods. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 17273–17283, 2021.

[14] Bryn Elesedy and Sheheryar Zaidi. Provably strict generalisation benefit for equivariant models. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, Proceedings of Machine Learning Research, pages 2959–2969. PMLR, 2021.

[15] Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, Proceedings of Machine Learning Research, pages 3165–3176. PMLR, 2020.

[16] Fabian Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[17] Jeong Min Jeon, Byeong U Park, and Ingrid Van Keilegom. Nonparametric regression on lie groups with measurement errors. *The Annals of Statistics*, 50(5):2973–3008, 2022.

[18] Jussi Klemelä. Estimation of densities and derivatives of densities with directional data. *J. Multivar. Anal.*, 73(1):18–40, April 2000.

[19] Anthony W. Knapp. *Lie Groups Beyond an Introduction*. Progress in Mathematics. Birkhäuser, 2nd edition, 2002.

[20] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, Proceedings of Machine Learning Research, pages 2752–2760. PMLR, 2018.

[21] Dima Kozakov, David R. Hall, Bing Xia, Kathryn A. Porter, Dzmitry Padhorny, Christine Yueh, Dmitri Beglov, and Sandor Vajda. The ClusPro web server for protein–protein docking. *Nature Protocols*, 12(2):255–278, 2017.

[22] Yi Lin and Hao Helen Zhang. Component selection and smoothing in multivariate nonparametric regression. *The Annals of Statistics*, 34(5), October 2006.

[23] Wilhelm Maier and Alfred Saupe. Eine einfache molekular-statistische Theorie der nematischen kristallinflüssigen Phase. Teil I. *Zeitschrift für Naturforschung A*, 14:882–889, 1959.

[24] Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, Proceedings of Machine Learning Research, pages 4363–4371. PMLR, 2019.

[25] Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Learning with invariances in random features and kernel models. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, Proceedings of Machine Learning Research, pages 3351–3418. PMLR, 2021.

[26] Garvesh Raskutti, Martin J. Wainwright, and Bin Yu. Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.*, 13:389–427, 2012.

[27] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, Proceedings of Machine Learning Research, pages 9323–9332. PMLR, 2021.

[28] Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053, 1982.

[29] Charles J Stone. Additive regression and other nonparametric models. *The annals of Statistics*, pages 689–705, 1985.

[30] Behrooz Tahmasebi and Stefanie Jegelka. The exact sample complexity gain from invariances for kernel regression. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

[31] Nathaniel Thomas, Tess E. Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *CoRR*, abs/1802.08219, 2018.

[32] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer series in statistics. Springer, 2009.

[33] Dmitri A. Varshalovich, Anatoly N. Moskalev, and Valeri K. Khersonskii. *Quantum Theory of Angular Momentum*. World Scientific, 1988.

[34] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

[35] Grace Wahba. *Spline Models for Observational Data*. SIAM, 1990.

[36] Maurice Weiler and Gabriele Cesa. General e(2)-equivariant steerable cnns. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 14334–14345, 2019.

[37] Yuhong Yang and Andrew Barron. Information-theoretic determination of minimax rates of convergence. *The Annals of Statistics*, 27(5):1564–1599, 1999.

A Notation and Background

This section collects the standard Peter–Weyl harmonic-analysis background used throughout the appendix.

Throughout the appendix, $\mathrm{SO}(3)$ is equipped with its normalized Haar measure μ. We write D^ℓ for the Wigner irreducible representation of degree $\ell = 0, 1, 2, \ldots$, with dimension $d_\ell = 2\ell + 1$ and Casimir eigenvalue $\lambda_\ell = \ell(\ell + 1)$; see [33, 19] for standard references. For the product group $\mathbf{G} = \mathrm{SO}(3)^n$, we write $\mathbf{g} = (g_1, \ldots, g_n)$ and use the product Haar measure $\mu^{\otimes n}$. Irreducible representations of $\mathbf{G}$ are indexed by $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n) \in \mathbb{Z}_{\geq 0}^n$, with dimension $d_{\boldsymbol{\ell}} = \prod_i (2\ell_i + 1)$ and Casimir eigenvalue $\lambda_{\boldsymbol{\ell}} = \sum_i \ell_i(\ell_i + 1)$.

Definition 4 (Sobolev norm)**.** $\|f\|_{H^s(\mathbf{G})}^2 := \sum_{\boldsymbol{\ell}} (1 + \lambda_{\boldsymbol{\ell}})^s \, d_{\boldsymbol{\ell}} \, \|\widehat{f}(\boldsymbol{\ell})\|_{\mathrm{HS}}^2$. *The same definition is used on* $\mathrm{SO}(3)^m$ *for any* m*, with* $\boldsymbol{\ell} \in \mathbb{Z}_{\geq 0}^m$ *and* $\lambda_{\boldsymbol{\ell}} = \sum_{i=1}^m \ell_i(\ell_i + 1)$.

Lemma 5 (Weyl counting)**.** *The spectral mode count on* $\mathrm{SO}(3)^m$ *(total* L^2*-orthogonal basis functions with Casimir* $\leq \Lambda$*; each irrep* $\boldsymbol{\ell}$ *contributes* $d_{\boldsymbol{\ell}}^2$ *basis functions) satisfies*

$$\mathsf{N}_m(\Lambda) := \sum_{\boldsymbol{\ell}: \lambda_{\boldsymbol{\ell}} \leq \Lambda} d_{\boldsymbol{\ell}}^2 \sim C_m \Lambda^{3m/2} \qquad \text{as } \Lambda \to \infty.$$

(We reserve N_m *for the mode count to avoid confusion with the sample size* N*.)*

Proof. On a single $\mathrm{SO}(3)$: $\sum_{\ell:\ell(\ell+1)\leq\Lambda} (2\ell + 1)^2 \sim \frac{4}{3}\Lambda^{3/2}$ by Euler–Maclaurin summation (each ℓ contributes $(2\ell + 1)^2 \sim 4\ell^2$ and $\ell_{\max} \sim \Lambda^{1/2}$). On the product $\mathrm{SO}(3)^m$: by the tensor product structure, $\mathsf{N}_m(\Lambda) = \sum_{\boldsymbol{\ell}:\lambda_{\boldsymbol{\ell}}\leq\Lambda} \prod (2\ell_i + 1)^2$. A standard lattice-point counting argument (see e.g. [1]) gives $\mathsf{N}_m(\Lambda) \sim C_m \Lambda^{3m/2}$. □

Lemma 6 (Sobolev embedding)**.** *If* $s > 3m/2$*:* $\|f\|_\infty \leq C_{\mathrm{emb}}(s, m)\|f\|_{H^s}$*, where* $C_{\mathrm{emb}}(s, m)^2 = \sum_{\boldsymbol{\ell}} d_{\boldsymbol{\ell}}^2 (1 + \lambda_{\boldsymbol{\ell}})^{-s}$.

Proof. By Cauchy–Schwarz on the Peter–Weyl expansion: $|f(\mathbf{g})| = |\sum_{\boldsymbol{\ell}} d_{\boldsymbol{\ell}} \mathrm{tr}(\widehat{f}(\boldsymbol{\ell}) D^{\boldsymbol{\ell}}(\mathbf{g}))| \leq \sum_{\boldsymbol{\ell}} d_{\boldsymbol{\ell}}^{3/2} \|\widehat{f}(\boldsymbol{\ell})\|_{\mathrm{HS}}$ (since $\|D^{\boldsymbol{\ell}}(\mathbf{g})\|_{\mathrm{HS}} = d_{\boldsymbol{\ell}}^{1/2}$). Applying Cauchy–Schwarz between the factors $(1 + \lambda_{\boldsymbol{\ell}})^{-s/2} d_{\boldsymbol{\ell}}$ and $(1 + \lambda_{\boldsymbol{\ell}})^{s/2} d_{\boldsymbol{\ell}}^{1/2} \|\widehat{f}(\boldsymbol{\ell})\|_{\mathrm{HS}}$ gives the result. By Lemma 5, the dyadic shell $\{\boldsymbol{\ell} : \Lambda \leq \lambda_{\boldsymbol{\ell}} < 2\Lambda\}$ contributes $\asymp \Lambda^{3m/2-s}$ to $C_{\mathrm{emb}}(s, m)^2$; summing over dyadic scales gives a geometric series, which converges iff $s > 3m/2$. □

Lemma 7 (Spectral truncation)**.** $\|f - f_\Lambda\|_{L^2}^2 \leq (1 + \Lambda)^{-s} \|f\|_{H^s}^2$*, where* f_Λ *retains all modes with* $\lambda_{\boldsymbol{\ell}} \leq \Lambda$.

Proof. $\|f - f_\Lambda\|_{L^2}^2 = \sum_{\lambda_{\boldsymbol{\ell}} > \Lambda} d_{\boldsymbol{\ell}} \|\widehat{f}(\boldsymbol{\ell})\|_{\mathrm{HS}}^2 \leq (1 + \Lambda)^{-s} \sum_{\lambda_{\boldsymbol{\ell}} > \Lambda} (1 + \lambda_{\boldsymbol{\ell}})^s d_{\boldsymbol{\ell}} \|\widehat{f}(\boldsymbol{\ell})\|_{\mathrm{HS}}^2 \leq (1 + \Lambda)^{-s} \|f\|_{H^s}^2$. □

B ANOVA Decomposition and Invariance

This section formalizes the ANOVA decomposition invoked in Section 3.1, distinguishing the general ANOVA subspace $\mathcal{W}_S$ (used for $\mathcal{F}_K$) from the invariance-restricted version $\mathcal{V}_S$ (used for $\mathcal{F}_{K,\mathrm{inv}}$).

Define $\mathrm{supp}(\boldsymbol{\ell}) := \{i : \ell_i \geq 1\}$. The ANOVA component at subset S is $f_S = \sum_{\boldsymbol{\ell}:\mathrm{supp}(\boldsymbol{\ell})=S} d_{\boldsymbol{\ell}} \, \mathrm{tr}(\widehat{f}(\boldsymbol{\ell}) D^{\boldsymbol{\ell}}(\mathbf{g}))$.

Definition 8 (Pure-S invariant ANOVA subspace $\mathcal{V}_S$)**.** *For* $|S| \geq 2$*,* $\mathcal{V}_S \subset L^2(\mathbf{G})$ *consists of functions satisfying: (a)* $\widehat{f}(\boldsymbol{\ell}) \neq 0 \Rightarrow \mathrm{supp}(\boldsymbol{\ell}) = S$*; (b)* $f(h\mathbf{g}) = f(\mathbf{g})$ *for all* $h \in \mathrm{SO}(3)$ *(diagonal invariance). For* $|S| = 0$*:* $\mathcal{V}_\emptyset = \{$*constants*$\}$*. For* $|S| = 1$*:* $\mathcal{V}_{\{i\}} = \{0\}$ *(an invariant function of* g_i *alone must be constant, with support* $\emptyset$ *not* $\{i\}$*).*

Definition 9 (General ANOVA subspace $\mathcal{W}_S$)**.** *For any* $S \subseteq [n]$*,* $\mathcal{W}_S := \{f \in L^2(\mathbf{G}) : \widehat{f}(\boldsymbol{\ell}) \neq 0 \Rightarrow \mathrm{supp}(\boldsymbol{\ell}) = S\}$*. Note* $\mathcal{V}_S \subset \mathcal{W}_S$ *(with the additional invariance constraint).*

Proposition 10 (General ANOVA orthogonality)**.** *$\mathcal{W}_S \perp \mathcal{W}_{S'}$ in $L^2(\mathbf{G})$ for $S \neq S'$. If $f = \sum_S f_S$ with $f_S \in \mathcal{W}_S$: $\|f\|_{L^2}^2 = \sum_S \|f_S\|_{L^2}^2$ and $\|f\|_{H^s}^2 = \sum_S \|f_S\|_{H^s}^2$.*

Proof. If $f \in \mathcal{W}_S$ and $g \in \mathcal{W}_{S'}$ with $S \neq S'$, their Peter–Weyl supports lie on disjoint index sets $\{\boldsymbol{\ell} : \mathrm{supp}(\boldsymbol{\ell}) = S\}$ and $\{\boldsymbol{\ell} : \mathrm{supp}(\boldsymbol{\ell}) = S'\}$. By Plancherel, all cross-inner-products vanish: $\langle f, g\rangle_{L^2} = 0$. The L^2 and H^s decompositions follow because the Sobolev weight $(1 + \lambda_{\boldsymbol{\ell}})^s$ depends only on $\boldsymbol{\ell}$, so the H^s-inner-product also vanishes on disjoint Fourier blocks. □

Proposition 11 (Invariant ANOVA orthogonality)**.** *$\mathcal{V}_S \perp \mathcal{V}_{S'}$ for $S \neq S'$. If $f = \sum_S f_S$ with $f_S \in \mathcal{V}_S$: $\|f\|_{L^2}^2 = \sum_S \|f_S\|_{L^2}^2$ and $\|f\|_{H^s}^2 = \sum_S \|f_S\|_{H^s}^2$.*

Proof. Follows from Proposition 10 since $\mathcal{V}_S \subset \mathcal{W}_S$. □

Lemma 12 (Invariance of ANOVA components)**.** *If f is diagonally invariant, then each ANOVA component f_S is also diagonally invariant.*

Proof. Let P_S denote the L^2-orthogonal projection onto the Fourier block with support exactly S, so that $f_S = P_S f$ (Definition 9). Let ρ_h denote the induced action of $h \in \mathrm{SO}(3)$ on functions, $(\rho_h f)(\mathbf{g}) := f(h\mathbf{g})$. The diagonal left action acts within each Peter–Weyl block $\mathcal{H}_{\boldsymbol{\ell}}$ and leaves the support $\mathrm{supp}(\boldsymbol{\ell})$ unchanged. Hence P_S commutes with ρ_h:

$$P_S(\rho_h f) = \rho_h(P_S f).$$

If f is diagonally invariant, i.e., $\rho_h f = f$ for all h, then

$$\rho_h f_S = \rho_h(P_S f) = P_S(\rho_h f) = P_S f = f_S,$$

so f_S is diagonally invariant. □

C Relative-Coordinate Pullback

This section develops the pullback operator $T_S : L^2(\mathrm{SO}(3)^{k-1}) \to L^2(\mathbf{G})$ that realizes the main-text identification $f_S(\mathbf{g}) = \varphi_S(g_{i_1}^{-1} g_{i_2}, \ldots, g_{i_1}^{-1} g_{i_k})$ for invariant k-body components. A direct pullback $T_S\varphi$ can contain lower-order invariant ANOVA leakage. For the full-support inputs used downstream, this leakage is captured precisely by the diagonal-invariant part of φ. The main result of this section (Theorem 18) shows that after projecting away this diagonal-invariant part on the relative-coordinate domain, the pure pullback $T_S^{\mathrm{pure}} := T_S \circ P_\Delta^\perp$ lands in the pure-S invariant ANOVA subspace $\mathcal{V}_S$. This is the construction used in the lower-bound packing (Appendix G) and is the source of the $3(k-1)$ effective dimension for invariant k-body components.

Definition 13 (Pullback)**.** *For $S = \{i_1, \ldots, i_k\}$, $|S| = k \geq 2$: $(T_S\varphi)(\mathbf{g}) = \varphi(g_{i_1}^{-1} g_{i_2}, \ldots, g_{i_1}^{-1} g_{i_k})$.*

Lemma 14 (L^2-isometry)**.** *T_S is a linear L^2-isometry: for all $\varphi, \varphi' \in L^2(\mathrm{SO}(3)^{k-1})$,*

$$\|T_S\varphi\|_{L^2(\mathbf{G})} = \|\varphi\|_{L^2(\mathrm{SO}(3)^{k-1})}, \qquad \langle T_S\varphi, T_S\varphi'\rangle_{L^2(\mathbf{G})} = \langle \varphi, \varphi'\rangle_{L^2(\mathrm{SO}(3)^{k-1})}.$$

Proof. Change variables $r_j = g_{i_1}^{-1} g_{i_j}$, using left-invariance of Haar measure. The integral over g_{i_1} and all g_j with $j \notin S$ yields factor 1, leaving the integral over the relative coordinates $(r_2, \ldots, r_k)$ on $\mathrm{SO}(3)^{k-1}$. The same change of variables applies to $\int |T_S\varphi|^2 \, d\mu^{\otimes n}$ and to $\int \overline{T_S\varphi}\, T_S\varphi' \, d\mu^{\otimes n}$, yielding both identities simultaneously. □

Lemma 15 (Invariance)**.** *$T_S\varphi$ is diagonally invariant: $(T_S\varphi)(h\mathbf{g}) = (T_S\varphi)(\mathbf{g})$ for all h.*

Proof. $(T_S\varphi)(h\mathbf{g}) = \varphi((hg_{i_1})^{-1}(hg_{i_2}), \ldots, (hg_{i_1})^{-1}(hg_{i_k})) = \varphi(g_{i_1}^{-1} h^{-1} h g_{i_2}, \ldots) = \varphi(g_{i_1}^{-1} g_{i_2}, \ldots) = (T_S\varphi)(\mathbf{g})$. □

Lemma 16 (Reference marginal)**.** *$\int (T_S\varphi)\, d\mu(g_{i_1}) = (A_\Delta\varphi)(g_{i_2}, \ldots, g_{i_k})$, where $A_\Delta\varphi(r_2, \ldots, r_k) := \int \varphi(hr_2, \ldots, hr_k)\, d\mu(h)$ is the averaging operator over the diagonal $\mathrm{SO}(3)$-action on $\mathrm{SO}(3)^{k-1}$. A_Δ is an orthogonal projection onto the invariant subspace, block-diagonal in Peter–Weyl, and satisfies $\|A_\Delta\varphi\|_{H^s} \leq \|\varphi\|_{H^s}$. The complementary projection is $P_\Delta^\perp := I - A_\Delta$, which is also an orthogonal projection (onto the diagonal-non-invariant subspace), block-diagonal in Peter–Weyl, and H^s-contractive: $\|P_\Delta^\perp\varphi\|_{H^s} \leq \|\varphi\|_{H^s}$.*

Proof. $\int (T_S\varphi)(g_{i_1},\dots)\,d\mu(g_{i_1}) = \int \varphi(g_{i_1}^{-1}g_{i_2},\dots,g_{i_1}^{-1}g_{i_k})\,d\mu(g_{i_1})$. Setting $h = g_{i_1}^{-1}$: this equals $\int \varphi(hg_{i_2},\dots,hg_{i_k})\,d\mu(h) = (A_\Delta\varphi)(g_{i_2},\dots,g_{i_k})$. In Peter–Weyl, A_Δ projects each block $\boldsymbol{\sigma}$ onto its diagonal-invariant subspace (the $\mathrm{SO}(3)$-isotypic component at the trivial representation), which is an orthogonal projection. Hence $\|A_\Delta\varphi\|_{H^s} \le \|\varphi\|_{H^s}$. Since $P_\Delta^\perp = I - A_\Delta$ is an orthogonal projection onto the L^2-complement of the range of A_Δ (also block-diagonal by the same Peter–Weyl structure), it also satisfies $\|P_\Delta^\perp\varphi\|_{H^s}^2 \le \|\varphi\|_{H^s}^2$; specifically, within each Peter–Weyl block (with constant Sobolev weight $(1+\lambda_{\boldsymbol{\sigma}})^s$), $P_\Delta^\perp$ is an orthogonal projection in L^2, hence norm-contractive. □

The next lemma quantifies, under the full spectral support assumption, when the pulled-back function $T_S\varphi$ lies cleanly in $\mathcal{V}_S$ versus when it "leaks" into lower-order $\mathcal{V}_{S'}$ subspaces: such leakage is captured precisely by the diagonal-invariant part of φ, and the $P_\Delta^\perp$-projected input pulls back cleanly.

Lemma 17 (Leakage decomposition)**.** *Let $\varphi \in L^2(\mathrm{SO}(3)^{k-1})$ have full spectral support (i.e., $\widehat{\varphi}(\boldsymbol{\sigma}) = 0$ whenever some $\sigma_j = 0$), and let $\Pi_{<S}$ be the L^2-orthogonal projection onto the subspace $\bigoplus_{S'\subsetneq S}\mathcal{V}_{S'}$. Then:*

(i) $T_S\varphi \in \bigoplus_{S'\subseteq S}\mathcal{V}_{S'}$.

(ii) $\Pi_{<S}(T_S\varphi) = \Pi_{<S}(T_S(A_\Delta\varphi))$.

(iii) $\Pi_{<S}(T_S(P_\Delta^\perp\varphi)) = 0$, *i.e.,* $T_S(P_\Delta^\perp\varphi) \in \mathcal{V}_S$.

Proof. (i): $(T_S\varphi)(\mathbf{g}) = \varphi(g_{i_1}^{-1}g_{i_2},\dots,g_{i_1}^{-1}g_{i_k})$ depends only on $\{g_i : i \in S\}$ and is diagonally invariant (Lemma 15). Every diagonally invariant function supported on S-coordinates lies in $\bigoplus_{S'\subseteq S}\mathcal{V}_{S'}$ by the ANOVA decomposition restricted to variables in S together with Lemma 12 (which ensures each ANOVA component is also diagonally invariant, hence in $\mathcal{V}_{S'}$). Applied to $P_\Delta^\perp\varphi \in L^2(\mathrm{SO}(3)^{k-1})$, the same conclusion gives $T_S(P_\Delta^\perp\varphi) \in \bigoplus_{S'\subseteq S}\mathcal{V}_{S'}$ as well.

(iii): Given (i), to prove $T_S(P_\Delta^\perp\varphi) \in \mathcal{V}_S$ it suffices to rule out all lower-order components, i.e., to show $T_S(P_\Delta^\perp\varphi) \perp \mathcal{V}_{S'}$ for every $S' \subsetneq S$. We split into two cases based on whether $i_1 \in S'$.

Case A: $i_1 \in S'$. Any $h \in \mathcal{V}_{S'}$, being diagonally invariant and depending only on $\{g_i : i \in S'\}$, can be written as $h(\mathbf{g}) = \hat{h}(g_{i_1}^{-1}g_{i_j} : i_j \in S' \setminus \{i_1\})$, a function of only $|S'|-1 < k-1$ of the S-relative coordinates. In the relative-coordinate representation on $\mathrm{SO}(3)^{k-1}$, $\hat{h}$ has at least one trivial Peter–Weyl factor ($\sigma_{j^*} = 0$ for an unused coordinate). Since $P_\Delta^\perp\varphi$ has full support (all $\sigma_j \ge 1$), Peter–Weyl orthogonality gives $\langle P_\Delta^\perp\varphi, \hat{h}\rangle_{L^2(\mathrm{SO}(3)^{k-1})} = 0$, and by Lemma 14, $\langle T_S(P_\Delta^\perp\varphi), h\rangle_{L^2(\mathbf{G})} = 0$.

Case B: $i_1 \notin S'$. Any $h \in \mathcal{V}_{S'}$ can be written as $h(\mathbf{g}) = \tilde{h}(g_{i_j}^{-1}g_{i_{j'}} : i_j, i_{j'} \in S')$ (relative coordinates internal to S'). In S-relative coordinates, $g_{i_j}^{-1}g_{i_{j'}} = r_j^{-1}r_{j'}$, so h becomes a function $H(r_2,\dots,r_k)$ that is diagonally invariant under $r_j \mapsto ur_j$. Therefore $H \in \mathrm{range}(A_\Delta)$ on $\mathrm{SO}(3)^{k-1}$. Since $P_\Delta^\perp\varphi \in \mathrm{range}(P_\Delta^\perp) = \mathrm{range}(A_\Delta)^\perp$, we have $\langle P_\Delta^\perp\varphi, H\rangle_{L^2(\mathrm{SO}(3)^{k-1})} = 0$, and again by Lemma 14, $\langle T_S(P_\Delta^\perp\varphi), h\rangle_{L^2(\mathbf{G})} = 0$.

Combining both cases, $T_S(P_\Delta^\perp\varphi) \perp \mathcal{V}_{S'}$ for every $S' \subsetneq S$, proving (iii).

(ii): By linearity of T_S, $T_S\varphi = T_S(A_\Delta\varphi) + T_S(P_\Delta^\perp\varphi)$. Applying $\Pi_{<S}$ and using (iii) to cancel the second term, $\Pi_{<S}(T_S\varphi) = \Pi_{<S}(T_S(A_\Delta\varphi))$. □

Combining the previous lemmas yields the main result of this section:

Theorem 18 (Pure-S pullback)**.** *Let $\varphi \in L^2(\mathrm{SO}(3)^{k-1})$ have full spectral support, i.e., $\widehat{\varphi}(\boldsymbol{\sigma}) = 0$ whenever some coordinate of $\boldsymbol{\sigma}$ is zero. Then*

$$T_S^{\mathrm{pure}}\varphi := T_S(P_\Delta^\perp\varphi) \in \mathcal{V}_S, \qquad \|T_S^{\mathrm{pure}}\varphi\|_{L^2} = \|P_\Delta^\perp\varphi\|_{L^2} \le \|\varphi\|_{L^2}.$$

For $k = 2$, $P_\Delta^\perp$ acts as the identity on full-support inputs, since $A_\Delta\varphi$ equals the constant Fourier mode $\widehat{\varphi}(0)$, which vanishes under the full-support condition.

Proof. Since $T_S(P_\Delta^\perp\varphi)$ depends only on the coordinates in S (by construction of T_S) and is diagonally invariant (Lemma 15), its ANOVA decomposition involves only subsets $S' \subseteq S$, and by

Lemma 12 each component is invariant. That is, $T_S(P_\Delta^\perp \varphi) = \sum_{S' \subseteq S} h_{S'}$ with $h_{S'} \in \mathcal{V}_{S'}$. By Lemma 17(iii), all lower-order components vanish: $h_{S'} = 0$ for $S' \subsetneq S$. Therefore only the S-component remains: $T_S(P_\Delta^\perp \varphi) = h_S \in \mathcal{V}_S$. L^2-norm: $\|T_S^{\mathrm{pure}} \varphi\|_{L^2} = \|T_S(P_\Delta^\perp \varphi)\|_{L^2} = \|P_\Delta^\perp \varphi\|_{L^2}$ by Lemma 14, and $\|P_\Delta^\perp \varphi\|_{L^2} \le \|\varphi\|_{L^2}$ by contractivity of orthogonal projections. □

Lemma 19 (Mode count)**.** *Let $k \ge 2$. There exist constants $\Lambda_{\min,k} > 0$ and $c_0, C_0 > 0$ (depending on k only) such that $\mathcal{U}_{k,\Lambda}^{\mathrm{pure}}$ is nonempty for all $\Lambda \ge \Lambda_{\min,k}$, and*

$$c_0\, \Lambda^{3(k-1)/2} \;\le\; m^{\mathrm{pure}}(\Lambda) \;:=\; \dim \mathcal{U}_{k,\Lambda}^{\mathrm{pure}} \;\le\; C_0\, \Lambda^{3(k-1)/2} \qquad \textit{for all } \Lambda \ge \Lambda_{\min,k}.$$

Equivalently, $m^{\mathrm{pure}}(\Lambda) \asymp \Lambda^{3(k-1)/2}$ as $\Lambda \to \infty$.

Proof. Define three subspaces of $L^2(\mathrm{SO}(3)^{k-1})$:

- $\mathcal{U}_{k,\Lambda}^{\mathrm{full}} := \mathrm{span}\{$Peter–Weyl modes $\boldsymbol{\sigma} = (\sigma_2, \dots, \sigma_k)$ on $\mathrm{SO}(3)^{k-1}$ with all $\sigma_j \ge 1$ and $\lambda_{\boldsymbol{\sigma}} \le \Lambda\}$ (full-support, spectrally truncated).
- $\mathcal{U}_{k,\Lambda}^{\Delta} := \mathcal{U}_{k,\Lambda}^{\mathrm{full}} \cap \mathrm{range}(A_\Delta)$ (the diagonal-invariant part).
- $\mathcal{U}_{k,\Lambda}^{\mathrm{pure}} := \mathcal{U}_{k,\Lambda}^{\mathrm{full}} \cap \mathrm{range}(P_\Delta^\perp)$ (the pure part used in the packing).

Step 1: Weyl counting (Lemma 5) gives $\mathsf{N}_{k-1}(\Lambda) \asymp \Lambda^{3(k-1)/2}$. The full-support subset (all $\ell_j \ge 1$) is this total minus the modes with some $\ell_j = 0$; the latter effectively live on $\mathrm{SO}(3)^{k-2}$ and contribute $O(\Lambda^{3(k-2)/2})$. Hence $\dim \mathcal{U}_{k,\Lambda}^{\mathrm{full}} \asymp \Lambda^{3(k-1)/2}$.

Step 2: $\dim \mathcal{U}_{k,\Lambda}^{\Delta} = O(\Lambda^{3(k-2)/2})$. The operator A_Δ projects onto functions on $\mathrm{SO}(3)^{k-1}$ that are invariant under the free diagonal left action of $\mathrm{SO}(3)$. Its range is therefore naturally identified with L^2 on the quotient

$$\mathrm{SO}(3)^{k-1} / \Delta_{\mathrm{SO}(3)} \cong \mathrm{SO}(3)^{k-2},$$

a compact manifold of real dimension $3(k-2)$. The restriction of the ambient product Casimir to this invariant subspace is an elliptic positive operator on the quotient, with principal symbol comparable to the quotient Laplace–Beltrami operator; its spectral counting function therefore has the same Weyl exponent as the Laplacian on a $3(k-2)$-dimensional compact manifold, giving $\dim \mathcal{U}_{k,\Lambda}^{\Delta} = O(\Lambda^{3(k-2)/2})$.

Step 3: Since A_Δ preserves each Peter–Weyl block $\boldsymbol{\sigma}$ (it acts block-diagonally in Peter–Weyl by Lemma 16) and hence preserves the Casimir cutoff, the truncated space $\mathcal{U}_{k,\Lambda}^{\mathrm{full}}$ is invariant under both A_Δ and $P_\Delta^\perp = I - A_\Delta$. Therefore

$$\mathcal{U}_{k,\Lambda}^{\mathrm{full}} = \mathcal{U}_{k,\Lambda}^{\Delta} \oplus \mathcal{U}_{k,\Lambda}^{\mathrm{pure}}$$

as an internal orthogonal direct sum.

Upper bound. From $\mathcal{U}_{k,\Lambda}^{\mathrm{pure}} \subseteq \mathcal{U}_{k,\Lambda}^{\mathrm{full}}$ and Step 1:

$$\dim \mathcal{U}_{k,\Lambda}^{\mathrm{pure}} \le \dim \mathcal{U}_{k,\Lambda}^{\mathrm{full}} \le C \Lambda^{3(k-1)/2}.$$

Lower bound. From the direct-sum decomposition and Steps 1–2:

$$\dim \mathcal{U}_{k,\Lambda}^{\mathrm{pure}} = \dim \mathcal{U}_{k,\Lambda}^{\mathrm{full}} - \dim \mathcal{U}_{k,\Lambda}^{\Delta} \ge C \Lambda^{3(k-1)/2} - C' \Lambda^{3(k-2)/2} \asymp \Lambda^{3(k-1)/2},$$

where the leading term dominates since $3(k-1)/2 > 3(k-2)/2$.

Explicitly, choose $\Lambda_{\min,k} := \max\{1, (2C'/C)^2\}$; then for all $\Lambda \ge \Lambda_{\min,k}$, $C' \Lambda^{3(k-2)/2} \le (C/2) \Lambda^{3(k-1)/2}$, giving $\dim \mathcal{U}_{k,\Lambda}^{\mathrm{pure}} \ge (C/2) \Lambda^{3(k-1)/2}$. Setting $c_0 := C/2$ (from this lower bound) and $C_0 := C$ (from Step 1's upper bound) yields the claimed bilateral bound. Nonemptiness of $\mathcal{U}_{k,\Lambda}^{\mathrm{pure}}$ at $\Lambda = \Lambda_{\min,k}$ follows from $m^{\mathrm{pure}}(\Lambda_{\min,k}) \ge c_0 \Lambda_{\min,k}^{3(k-1)/2} > 0$. □

D Sobolev Norm Equivalence

The Sobolev norm equivalence below ensures that the reduced-domain analysis on $\mathrm{SO}(3)^{k-1}$ matches the original Sobolev-ball constraint on $\mathbf{G}$. Theorem 20 establishes bilateral bounds between $\|T_S\varphi\|_{H^s(\mathbf{G})}$ and $\|\varphi\|_{H^s(\mathrm{SO}(3)^{k-1})}$, with constants depending only on k. Corollary 21 extends this to an *existence* result: for the chosen reference coordinate i_1, every invariant k-body component $f_S \in \mathcal{V}_S$ admits an L^2-unique pure relative-coordinate representative $\varphi_S^{\mathrm{pure}}$ with $\|\varphi_S^{\mathrm{pure}}\|_{H^s} \le \|f_S\|_{H^s}$. Together, they yield a global budget $\sum_{k=2}^{K}\sum_{|S|=k}\|\varphi_S^{\mathrm{pure}}\|_{H^s}^2 \le R^2$ that is essential for the upper bound of Appendix F.

Theorem 20 (Sobolev norm equivalence under pullback)**.** *Let $S = \{i_1, \dots, i_k\} \subseteq [n]$ with $|S| = k \ge 2$, and let $\varphi \in H^s(\mathrm{SO}(3)^{k-1})$. Since $T_S\varphi$ depends only on coordinates in S, we identify it with the corresponding function on $\mathrm{SO}(3)^k$ (indices in S), and have*

$$\|\varphi\|_{H^s(\mathrm{SO}(3)^{k-1})} \;\le\; \|T_S\varphi\|_{H^s(\mathrm{SO}(3)^k)} \;\le\; k^{s/2}\,\|\varphi\|_{H^s(\mathrm{SO}(3)^{k-1})}.$$

When $T_S\varphi$ is viewed as a function on $\mathbf{G} = \mathrm{SO}(3)^n$ (via the trivial Fourier mode $\ell_i = 0$ on coordinates $i \notin S$, so $T_S\varphi$ does not depend on these coordinates), its $H^s(\mathbf{G})$ norm coincides with its $H^s(\mathrm{SO}(3)^k)$ norm:

$$\|T_S\varphi\|_{H^s(\mathbf{G})} = \|T_S\varphi\|_{H^s(\mathrm{SO}(3)^k)}.$$

Proof. We first prove the bilateral bound on $\mathrm{SO}(3)^k$; the embedding into $\mathbf{G}$ is addressed at the end.

Proof for $k = 2$. By Lemma 14, $\|T_S\varphi\|_{L^2} = \|\varphi\|_{L^2}$. Using $D^\ell(g_1^{-1}g_2) = D^\ell(g_1)^* D^\ell(g_2)$, the pullback maps a Peter–Weyl mode ℓ on $\mathrm{SO}(3)$ to the block (ℓ,ℓ) on $\mathrm{SO}(3)^2$ with Casimir $2\lambda_\ell$. Since $(1+\lambda_\ell) \le (1+2\lambda_\ell) \le 2(1+\lambda_\ell)$, we obtain

$$\|\varphi\|_{H^s} \le \|T_S\varphi\|_{H^s} \le 2^{s/2}\|\varphi\|_{H^s}.$$

Proof for general $k \ge 3$.

Step 1: Block structure. T_S maps the Peter–Weyl block $\mathcal{H}_{\boldsymbol{\sigma}}$ on $\mathrm{SO}(3)^{k-1}$ (with $\boldsymbol{\sigma} = (\sigma_2, \dots, \sigma_k)$) into the workspace

$$\mathcal{Y}_{\boldsymbol{\sigma}} \;=\; \bigoplus_{\ell_1 \in \mathrm{CG}(\boldsymbol{\sigma})} \mathcal{H}_{(\ell_1,\,\sigma_2,\dots,\sigma_k)}$$

on $\mathrm{SO}(3)^k$, where ℓ_1 ranges over the Clebsch–Gordan decomposition of $D^{\sigma_2} \otimes \cdots \otimes D^{\sigma_k}$. Crucially, the coordinate indices $j = 2, \dots, k$ retain their labels: $\ell_j = \sigma_j$ for these j. The workspaces $\mathcal{Y}_{\boldsymbol{\sigma}}$ and $\mathcal{Y}_{\boldsymbol{\sigma}'}$ are mutually orthogonal for $\boldsymbol{\sigma} \ne \boldsymbol{\sigma}'$ (distinct Fourier blocks on $\mathrm{SO}(3)^k$).

Step 2: Per-block L^2-isometry. Applying Lemma 14 to a single-block input $\varphi_{\boldsymbol{\sigma}} \in \mathcal{H}_{\boldsymbol{\sigma}}$:

$$\|T_S\varphi_{\boldsymbol{\sigma}}\|_{L^2(\mathrm{SO}(3)^k)}^2 = \|\varphi_{\boldsymbol{\sigma}}\|_{L^2(\mathrm{SO}(3)^{k-1})}^2 \quad \text{for each } \boldsymbol{\sigma}.$$

In Fourier coefficients, this reads

$$\sum_{\ell_1 \in \mathrm{CG}(\boldsymbol{\sigma})} d_{(\ell_1,\boldsymbol{\sigma})}\,\|\widehat{(T_S\varphi)}(\ell_1,\boldsymbol{\sigma})\|_{\mathrm{HS}}^2 \;=\; d_{\boldsymbol{\sigma}}\,\|\widehat{\varphi}(\boldsymbol{\sigma})\|_{\mathrm{HS}}^2. \tag{3}$$

Step 3: Casimir bound. From the Clebsch–Gordan range, $\ell_1 \le \sigma_2 + \cdots + \sigma_k$. By Cauchy–Schwarz, $(\sigma_2 + \cdots + \sigma_k)^2 \le (k-1)\sum_{j=2}^k \sigma_j^2$, so $\ell_1^2 \le (k-1)\sum_j \sigma_j^2$. Since $k \ge 2$, we also have $\ell_1 \le \sum_j \sigma_j \le (k-1)\sum_j \sigma_j$. Therefore

$$\lambda_{\ell_1} = \ell_1(\ell_1+1) = \ell_1^2 + \ell_1 \;\le\; (k-1)\sum_j \sigma_j^2 + (k-1)\sum_j \sigma_j \;=\; (k-1)\sum_j \lambda_{\sigma_j} \;=\; (k-1)\,\lambda_{\boldsymbol{\sigma}}.$$

Consequently,

$$\lambda_{\boldsymbol{\sigma}} \;\le\; \lambda_{\ell_1} + \lambda_{\boldsymbol{\sigma}} \;\le\; k\,\lambda_{\boldsymbol{\sigma}},$$

and equivalently $1 + \lambda_{\boldsymbol{\sigma}} \le 1 + \lambda_{\ell_1} + \lambda_{\boldsymbol{\sigma}} \le 1 + k\lambda_{\boldsymbol{\sigma}} \le k(1+\lambda_{\boldsymbol{\sigma}})$, using $k \ge 2$.

Step 4: Combine. The H^s-norm of $T_S\varphi$ on $\mathrm{SO}(3)^k$ expands over the image blocks:

$$\|T_S\varphi\|^2_{H^s(\mathrm{SO}(3)^k)} \;=\; \sum_{\boldsymbol{\sigma}} \sum_{\ell_1 \in \mathrm{CG}(\boldsymbol{\sigma})} (1+\lambda_{\ell_1}+\lambda_{\boldsymbol{\sigma}})^s \, d_{(\ell_1,\boldsymbol{\sigma})} \, \|\widehat{(T_S\varphi)}(\ell_1,\boldsymbol{\sigma})\|^2_{\mathrm{HS}}.$$

Upper bound. By Step 3, $(1+\lambda_{\ell_1}+\lambda_{\boldsymbol{\sigma}})^s \le k^s(1+\lambda_{\boldsymbol{\sigma}})^s$. Pulling the $\boldsymbol{\sigma}$-dependent weight outside the ℓ_1-sum and applying the per-block isometry (3):

$$\begin{aligned}\|T_S\varphi\|^2_{H^s} \;&\le\; k^s \sum_{\boldsymbol{\sigma}} (1+\lambda_{\boldsymbol{\sigma}})^s \sum_{\ell_1} d_{(\ell_1,\boldsymbol{\sigma})} \|\widehat{(T_S\varphi)}(\ell_1,\boldsymbol{\sigma})\|^2_{\mathrm{HS}} \\ &=\; k^s \sum_{\boldsymbol{\sigma}} (1+\lambda_{\boldsymbol{\sigma}})^s d_{\boldsymbol{\sigma}} \|\widehat{\varphi}(\boldsymbol{\sigma})\|^2_{\mathrm{HS}} \;=\; k^s \|\varphi\|^2_{H^s}.\end{aligned}$$

Lower bound. Since $(1+\lambda_{\ell_1}+\lambda_{\boldsymbol{\sigma}}) \ge (1+\lambda_{\boldsymbol{\sigma}})$, the same calculation gives

$$\|T_S\varphi\|^2_{H^s} \;\ge\; \sum_{\boldsymbol{\sigma}} (1+\lambda_{\boldsymbol{\sigma}})^s d_{\boldsymbol{\sigma}} \|\widehat{\varphi}(\boldsymbol{\sigma})\|^2_{\mathrm{HS}} \;=\; \|\varphi\|^2_{H^s}.$$

Embedding into G. When $T_S\varphi$ is viewed on $\mathbf{G}$, its Peter–Weyl expansion has $\widehat{(T_S\varphi)}(\boldsymbol{\ell}) \neq 0$ only for $\boldsymbol{\ell}$ with $\ell_i = 0$ for $i \notin S$. For such $\boldsymbol{\ell}$: the Casimir weight $(1+\lambda_{\boldsymbol{\ell}})^s$ on $\mathbf{G}$ equals the weight on $\mathrm{SO}(3)^k$ (since $\lambda_{\ell_i} = 0$ for $i \notin S$), and the dimension factor $d_{\boldsymbol{\ell}} = \prod_i (2\ell_i+1)$ reduces to $\prod_{i\in S}(2\ell_i+1) = d_{\boldsymbol{\ell}|_S}$ (the trivial factors contribute 1 each). Hence $\|T_S\varphi\|^2_{H^s(\mathbf{G})} = \|T_S\varphi\|^2_{H^s(\mathrm{SO}(3)^k)}$ as claimed. □

Corollary 21 (Pure relative-coordinate representative)**.** *Let $S = \{i_1,\dots,i_k\} \subseteq [n]$ with $|S| = k \ge 2$, and let $f_S \in \mathcal{V}_S$ with $\|f_S\|_{H^s(\mathbf{G})} < \infty$. Then there exists a representative $\varphi_S^{\mathrm{pure}} \in H^s(\mathrm{SO}(3)^{k-1})$, unique up to Haar-null sets, with full spectral support, lying in $\mathrm{range}(P_\Delta^\perp)$, such that*

$$f_S \;=\; T_S\varphi_S^{\mathrm{pure}}, \qquad \|\varphi_S^{\mathrm{pure}}\|_{H^s(\mathrm{SO}(3)^{k-1})} \;\le\; \|f_S\|_{H^s(\mathbf{G})}.$$

Consequently, if $f \;=\; f_0 + \sum_{k=2}^K \sum_{|S|=k} f_S \in \mathcal{F}_{K,\mathrm{inv}}$ with $\|f\|_{H^s(\mathbf{G})} \le R$, then

$$\sum_{k=2}^{K} \sum_{|S|=k} \|\varphi_S^{\mathrm{pure}}\|^2_{H^s} \;\le\; R^2.$$

Proof. Step A (Existence of some representative). The pullback $T_S : L^2(\mathrm{SO}(3)^{k-1}) \to L^2(\mathbf{G})$ is an isometric embedding (Lemma 14) whose image consists of all $L^2(\mathbf{G})$ functions that (i) depend only on $\{g_i : i \in S\}$ and (ii) are diagonally invariant. Indeed, if $F(\mathbf{g})$ satisfies (i) and (ii), the map $F \mapsto \varphi$ inverting the pullback is defined by the L^2-isometric change-of-variables of Lemma 14; for smooth representatives it admits the explicit pointwise form

$$\varphi(r_2,\dots,r_k) \;:=\; F(e,r_2,\dots,r_k) \quad \text{on } \mathrm{SO}(3)^{k-1},$$

and the same map is extended to general L^2 functions by density / by the isometric identity of Lemma 14. Diagonal invariance $F(hg_{i_1}, hg_{i_2}, \dots, hg_{i_k}) = F(g_{i_1},\dots,g_{i_k})$ with $h = g_{i_1}^{-1}$ yields

$$F(g_{i_1},\dots,g_{i_k}) = F(e, g_{i_1}^{-1}g_{i_2},\dots,g_{i_1}^{-1}g_{i_k}) = \varphi(g_{i_1}^{-1}g_{i_2},\dots,g_{i_1}^{-1}g_{i_k}) = (T_S\varphi)(\mathbf{g}),$$

so $F = T_S\varphi$. Since $\mathcal{V}_S$ is a closed subspace of the image of T_S, every $f_S \in \mathcal{V}_S$ has an L^2-unique preimage $\varphi_S \in L^2(\mathrm{SO}(3)^{k-1})$ satisfying $f_S = T_S\varphi_S$ (uniqueness up to Haar-null sets, from the L^2-isometry of T_S).

Step B (Full spectral support of φ_S). By Step 1 of Theorem 20, T_S maps a Peter–Weyl block $\boldsymbol{\sigma}$ on $\mathrm{SO}(3)^{k-1}$ into the workspace $\mathcal{Y}_{\boldsymbol{\sigma}}$ on $\mathrm{SO}(3)^k$, consisting of blocks $(\ell_1,\sigma_2,\dots,\sigma_k)$ with the constraint $\ell_j = \sigma_j$ for $j = 2,\dots,k$. Because the workspaces $\mathcal{Y}_{\boldsymbol{\sigma}}$ are mutually orthogonal for distinct $\boldsymbol{\sigma}$ (Theorem 20 Step 1), a contribution from one input block cannot be cancelled by contributions from other input blocks. Therefore, if $\widehat{\varphi}_S(\boldsymbol{\sigma}) \neq 0$ for some $\boldsymbol{\sigma}$ with $\sigma_{j^*} = 0$, then $T_S\varphi_S = f_S$ would have a nonzero Fourier mode with $\ell_{j^*} = 0$. But $f_S \in \mathcal{V}_S$ requires Fourier support exactly S (all

$\ell_i \geq 1$ for $i \in S$), a contradiction. Hence φ_S has full spectral support on $\mathrm{SO}(3)^{k-1}$: $\widehat{\varphi}_S(\boldsymbol{\sigma}) = 0$ for every $\boldsymbol{\sigma}$ with $\sigma_j = 0$ for some j.

Step C (Reduction to $P_\Delta^\perp$-range). Decompose $\varphi_S = A_\Delta \varphi_S + P_\Delta^\perp \varphi_S$ (Lemma 16). We argue that $T_S(A_\Delta \varphi_S)$ lies in a subspace of $L^2(\mathbf{G})$ orthogonal to $\mathcal{V}_S$; combined with $f_S \in \mathcal{V}_S$, this forces $T_S(A_\Delta \varphi_S) = 0$, yielding $f_S = T_S(P_\Delta^\perp \varphi_S)$.

Since φ_S has full spectral support (Step B), Theorem 18 applied to φ_S yields $T_S(P_\Delta^\perp \varphi_S) \in \mathcal{V}_S$ directly.

For the diagonal-invariant part $A_\Delta \varphi_S$, use its diagonal invariance on $\mathrm{SO}(3)^{k-1}$ (invoked with $h = g_{i_1}$):

$$(T_S(A_\Delta \varphi_S))(\mathbf{g}) = (A_\Delta \varphi_S)(g_{i_1}^{-1} g_{i_2}, \ldots, g_{i_1}^{-1} g_{i_k}) = (A_\Delta \varphi_S)(g_{i_2}, \ldots, g_{i_k}),$$

so $T_S(A_\Delta \varphi_S)$ depends only on $\{g_i : i \in S \setminus \{i_1\}\}$, hence has Fourier support contained in $S \setminus \{i_1\}$. As a result, $T_S(A_\Delta \varphi_S)$ lies in $\bigoplus_{S' \subseteq S \setminus \{i_1\}} \mathcal{V}_{S'}$, which is orthogonal to $\mathcal{V}_S$ (since each $\mathcal{V}_{S'}$ in this sum has Fourier support missing i_1, while $\mathcal{V}_S$ requires $\ell_{i_1} \geq 1$).

Combining: in the decomposition

$$f_S = T_S(A_\Delta \varphi_S) + T_S(P_\Delta^\perp \varphi_S),$$

both f_S and $T_S(P_\Delta^\perp \varphi_S)$ lie in $\mathcal{V}_S$, while $T_S(A_\Delta \varphi_S) \perp \mathcal{V}_S$. The only way this is consistent is $T_S(A_\Delta \varphi_S) = 0$, giving

$$f_S = T_S(P_\Delta^\perp \varphi_S).$$

Since T_S is an L^2-isometry (hence injective on $L^2(\mathrm{SO}(3)^{k-1})$), $T_S(A_\Delta \varphi_S) = 0$ implies $A_\Delta \varphi_S = 0$; equivalently, the preimage φ_S itself already lies in $\mathrm{range}(P_\Delta^\perp)$, so $\varphi_S = P_\Delta^\perp \varphi_S$. Set $\varphi_S^{\mathrm{pure}} := P_\Delta^\perp \varphi_S = \varphi_S$; by Step B, $\varphi_S^{\mathrm{pure}}$ has full spectral support.

Step D (Norm bound). By Theorem 20 (lower bound) and the H^s-contractivity of $P_\Delta^\perp$ (Lemma 16):

$$\|\varphi_S^{\mathrm{pure}}\|_{H^s} = \|P_\Delta^\perp \varphi_S\|_{H^s} \;\leq\; \|\varphi_S\|_{H^s} \;\leq\; \|T_S \varphi_S\|_{H^s} = \|f_S\|_{H^s}.$$

Step E (Global budget). By invariant ANOVA H^s-orthogonality (Proposition 11):

$$\sum_{k=2}^{K} \sum_{|S|=k} \|\varphi_S^{\mathrm{pure}}\|_{H^s}^2 \;\leq\; \sum_{k=2}^{K} \sum_{|S|=k} \|f_S\|_{H^s}^2 \;\leq\; \|f\|_{H^s}^2 \;\leq\; R^2.$$

□

E Sup-Norm Bound for $\mathcal{F}_{K,\mathrm{inv}}$

This lemma establishes a uniform sup-norm bound on $\mathcal{F}_{K,\mathrm{inv}}$, used in the variance analysis (Proposition 25) and the adaptive proof (Appendix J).

Lemma 22 (Sup-norm bound for $\mathcal{F}_{K,\mathrm{inv}}$)**.** *Assume $K \geq 2$ and $s > 3(K-1)/2$. If $f \in \mathcal{F}_{K,\mathrm{inv}}$ and $\|f\|_{H^s(\mathbf{G})} \leq R$, then*

$$\|f\|_\infty \;\leq\; C_{\mathrm{sup}}\, R, \qquad C_{\mathrm{sup}} \;:=\; \sqrt{\sum_{k=0}^{K} \binom{n}{k}} \cdot C_{\mathrm{emb}}(s, K-1).$$

Proof. Decompose $f = f_0 + \sum_{k=2}^{K} \sum_{|S|=k} f_S$, with no $k{=}1$ terms under invariance (main text, Section 3.1). The singleton count $\binom{n}{1}$ is nevertheless included in C_{sup} for a uniform combinatorial expression; the corresponding components vanish and contribute zero to the sum.

Constant component ($k = 0$). Since f_0 is constant on $\mathbf{G}$,

$$\|f_0\|_\infty \;=\; \|f_0\|_{L^2} \;=\; \|f_0\|_{H^s} \;\leq\; C_{\mathrm{emb}}(s, K-1)\, \|f_0\|_{H^s},$$

where the final inequality uses $C_{\mathrm{emb}}(s, K-1) \geq 1$ (the $\boldsymbol{\ell} = \mathbf{0}$ term contributes $d_{\mathbf{0}}^2 (1+0)^{-s} = 1$ to the defining series of C_{emb}^2, and all other terms are nonnegative).

k-body components with $|S| = k \geq 2$. By Corollary 21, $f_S = T_S \varphi_S^{\mathrm{pure}}$ with $\|\varphi_S^{\mathrm{pure}}\|_{H^s(\mathrm{SO}(3)^{k-1})} \leq \|f_S\|_{H^s(\mathbf{G})}$. The relative-coordinate map $z_S : \mathbf{G} \to \mathrm{SO}(3)^{k-1}$, $\mathbf{g} \mapsto (g_{i_1}^{-1} g_{i_2}, \ldots, g_{i_1}^{-1} g_{i_k})$, is surjective (for any target $(r_2, \ldots, r_k)$, set g_{i_1} arbitrary and $g_{i_j} = g_{i_1} r_j$), so $\|f_S\|_\infty = \|\varphi_S^{\mathrm{pure}}\|_\infty$. Applying the Sobolev embedding (Lemma 6) on $\mathrm{SO}(3)^{k-1}$ with $m = k-1$:

$$\|\varphi_S^{\mathrm{pure}}\|_\infty \;\leq\; C_{\mathrm{emb}}(s, k-1)\, \|\varphi_S^{\mathrm{pure}}\|_{H^s} \;\leq\; C_{\mathrm{emb}}(s, K-1)\, \|\varphi_S^{\mathrm{pure}}\|_{H^s},$$

where $C_{\mathrm{emb}}(s, m)$ is non-decreasing in m (the defining series on $\mathbb{Z}_{\geq 0}^m$ is embedded in the one on $\mathbb{Z}_{\geq 0}^{K-1}$ via the trivial modes). The embedding constant is finite because $s > 3(K-1)/2 \geq 3m/2$ for all $m \leq K-1$. Combining:

$$\|f_S\|_\infty \;\leq\; C_{\mathrm{emb}}(s, K-1)\, \|f_S\|_{H^s}.$$

Combining across subsets. Let $M := \sum_{k=0}^{K} \binom{n}{k}$ be the total number of subsets of $[n]$ with $|S| \leq K$ (including $\emptyset$). By the triangle inequality and Cauchy–Schwarz in $\mathbb{R}^M$:

$$\begin{aligned}\|f\|_\infty \;\leq\; \sum_{|S| \leq K} \|f_S\|_\infty \;&\leq\; C_{\mathrm{emb}}(s, K-1) \sum_{|S| \leq K} \|f_S\|_{H^s} \\ &\leq\; C_{\mathrm{emb}}(s, K-1) \sqrt{M} \sqrt{\sum_{|S| \leq K} \|f_S\|_{H^s}^2}.\end{aligned}$$

By invariant ANOVA H^s-orthogonality (Proposition 11), $\sum_{|S| \leq K} \|f_S\|_{H^s}^2 \leq \|f\|_{H^s}^2 \leq R^2$. Therefore

$$\|f\|_\infty \;\leq\; C_{\mathrm{emb}}(s, K-1) \sqrt{M}\, R \;=\; C_{\mathrm{sup}}\, R.$$

□

F Upper Bound for $\mathcal{F}_{K,\mathrm{inv}}$

This section proves the upper-bound half of Theorem 1 for $\mathcal{F}_{K,\mathrm{inv}}$ via spectral projection estimators on the relative-coordinate domain $\mathrm{SO}(3)^{k-1}$.

Basis convention. Throughout this section, we use a real L^2-orthonormal basis for each retained spectral block, obtained from the real and imaginary parts of the Wigner-D matrix coefficients. This convention is natural for real-valued regression targets ($Y_\alpha \in \mathbb{R}$) and makes the estimator coefficients real-valued; it eliminates the need for complex conjugation in the L^2 inner products. All norms on $\mathrm{SO}(3)^{k-1}$ and $\mathbf{G}$ are defined with respect to the normalized Haar measure.

Proof outline.

- The estimator's basis functions are pulled back via T_S into $\mathcal{V}_S$ (Theorem 18), so component estimators satisfy $\widehat{f}_S \in \mathcal{V}_S$ a.s., allowing pathwise orthogonal risk decomposition (Proposition 11).
- Unbiasedness (Proposition 24) uses invariant ANOVA orthogonality and the inner-product preservation of T_S (Lemma 14), extracting f_S^*'s pure representative from Corollary 21.
- Variance (Proposition 25) exploits the Haar-uniform distribution of the relative coordinate $z_S^{(\alpha)}$ and the uniform sup-norm bound (Lemma 22).
- The final rate (Theorem 26) uses *level-wise common cutoffs* $\Lambda_k \asymp \max\{\Lambda_{\min,k},\, (N/\binom{n}{k})^{2/(2s+3(k-1))}\}$ that depend only on the sample size N, the problem parameters (n, K, k, s), and known universal constants (R, σ)—not on unknown Sobolev budgets $\|\varphi_S^*\|_{H^s}$. Bias is analyzed directly on the relative domain $\mathrm{SO}(3)^{k-1}$ (via Lemma 7), matching the estimator's actual truncation. The global budget from Corollary 21 sums level-wise component risks; top body-order dominance at $k = K$ yields the rate.

Definition 23 (Estimator)**.** *Fix $K \geq 2$ and $s > 3(K-1)/2$. For each $|S| = k \in \{2, \ldots, K\}$:*

- ***Features.*** *Given sample $X_\alpha = (g_1^{(\alpha)}, \ldots, g_n^{(\alpha)}) \sim \mu^{\otimes n}$, form the S-relative coordinates*

$$z_S^{(\alpha)} \;=\; \big((g_{i_1}^{(\alpha)})^{-1} g_{i_2}^{(\alpha)}, \ldots, (g_{i_1}^{(\alpha)})^{-1} g_{i_k}^{(\alpha)}\big) \;\in\; \mathrm{SO}(3)^{k-1}.$$

- ***Basis.*** *Let $\{\phi_{S,j}\}_{j=1}^{M_S}$ be a real L^2-orthonormal basis of the full-support $P_\Delta^\perp$-range on $\mathrm{SO}(3)^{k-1}$ at relative Casimir level $\le \Lambda_k$, where $M_S \le C_1 \Lambda_k^{3(k-1)/2}$ by Lemma 19.*
- ***Pulled-back basis (elements of $\mathcal{V}_S$).*** *Each $\phi_{S,j}$ has full spectral support and lies in $\mathrm{range}(P_\Delta^\perp)$, so $T_S\phi_{S,j} \in \mathcal{V}_S$ by Theorem 18. By Lemma 14 (inner-product preservation), $\{T_S\phi_{S,j}\}_j$ is $L^2(\mathbf{G})$-orthonormal.*
- ***Coefficient.*** *Define*

$$\widehat{a}_{S,j} \;:=\; \frac{1}{N}\sum_{\alpha=1}^{N} Y_\alpha\, \phi_{S,j}(z_S^{(\alpha)}) \;=\; \frac{1}{N}\sum_{\alpha=1}^{N} Y_\alpha\, (T_S\phi_{S,j})(X_\alpha),$$

using $(T_S\phi_{S,j})(\mathbf{g}) = \phi_{S,j}(z_S(\mathbf{g}))$ for the second form.

- ***Component estimator.***

$$\widehat{f}_S \;:=\; \sum_{j=1}^{M_S} \widehat{a}_{S,j}\cdot T_S\phi_{S,j} \;\in\; \mathcal{V}_S \quad \textit{a.s.}$$

Constant term. $\widehat{f}_0 := \bar{Y} = \frac{1}{N}\sum_\alpha Y_\alpha$ *(a scalar, viewed as a constant function on $\mathbf{G}$; $\widehat{f}_0 \in \mathcal{V}_\emptyset$).*

Full estimator. $\widehat{f}_N := \widehat{f}_0 + \sum_{k=2}^{K}\sum_{|S|=k} \widehat{f}_S$.

The cutoffs Λ_k are specified in the proof of Theorem 26 and depend on N, n, K, s, R, σ only—no target-dependent oracle knowledge is used.

Proposition 24 (Unbiasedness)**.** *Let $\varphi_S^{\mathrm{pure}}$ denote the pure relative-coordinate representative of f_S^* (Corollary 21). Then for each retained mode,*

$$\mathbb{E}[\widehat{a}_{S,j}] \;=\; \langle \varphi_S^{\mathrm{pure}}, \phi_{S,j}\rangle_{L^2(\mathrm{SO}(3)^{k-1})}.$$

Proof. Under the observation model $Y_\alpha = f^*(X_\alpha) + \xi_\alpha$ with $\mathbb{E}[\xi_\alpha] = 0$ and $\xi_\alpha \perp\!\!\!\perp X_\alpha$, and using i.i.d. sampling:

$$\mathbb{E}[\widehat{a}_{S,j}] \;=\; \mathbb{E}\big[Y_1\cdot (T_S\phi_{S,j})(X_1)\big] \;=\; \mathbb{E}\big[f^*(X_1)\cdot (T_S\phi_{S,j})(X_1)\big] \;=\; \langle f^*,\, T_S\phi_{S,j}\rangle_{L^2(\mathbf{G})}.$$

Decompose $f^* = \sum_{S''} f^*_{S''}$ by invariant ANOVA (Appendix B):

$$\langle f^*,\, T_S\phi_{S,j}\rangle_{L^2(\mathbf{G})} \;=\; \sum_{S''}\langle f^*_{S''},\, T_S\phi_{S,j}\rangle_{L^2(\mathbf{G})}.$$

For $S'' \neq S$: $f^*_{S''} \in \mathcal{V}_{S''}$ and $T_S\phi_{S,j} \in \mathcal{V}_S$, so $\langle f^*_{S''}, T_S\phi_{S,j}\rangle_{L^2} = 0$ by Proposition 11.

For $S'' = S$: By Corollary 21, $f_S^* = T_S\varphi_S^{\mathrm{pure}}$. Applying the inner-product preservation of T_S (Lemma 14):

$$\langle f_S^*, T_S\phi_{S,j}\rangle_{L^2(\mathbf{G})} \;=\; \langle T_S\varphi_S^{\mathrm{pure}}, T_S\phi_{S,j}\rangle_{L^2(\mathbf{G})} \;=\; \langle \varphi_S^{\mathrm{pure}}, \phi_{S,j}\rangle_{L^2(\mathrm{SO}(3)^{k-1})}.$$

□

Proposition 25 (Variance)**.** *Assume $s > 3(K-1)/2$. For each retained mode,*

$$\mathrm{Var}(\widehat{a}_{S,j}) \;\le\; \frac{C_{\sup}^2 R^2 + \sigma^2}{N}.$$

Proof. Since the samples are i.i.d.,

$$\mathrm{Var}(\widehat{a}_{S,j}) \;=\; \frac{1}{N}\mathrm{Var}\big(Y_1\cdot\phi_{S,j}(z_S^{(1)})\big) \;\le\; \frac{1}{N}\mathbb{E}\big[Y_1^2\cdot\phi_{S,j}^2(z_S^{(1)})\big].$$

Expanding Y_1^2. With $Y_1 = f^*(X_1) + \xi_1$, $\mathbb{E}[\xi_1] = 0$, $\mathbb{E}[\xi_1^2] = \sigma^2$, and $\xi_1 \perp\!\!\!\perp X_1$:

$$\mathbb{E}[Y_1^2\phi_{S,j}^2(z_S^{(1)})] \;=\; \mathbb{E}[f^*(X_1)^2\phi_{S,j}^2(z_S^{(1)})] + \sigma^2\mathbb{E}[\phi_{S,j}^2(z_S^{(1)})],$$

the cross term $2\mathbb{E}[f^*(X_1)\xi_1\phi^2_{S,j}(z_S^{(1)})]$ vanishes since $\mathbb{E}[\xi_1 \mid X_1] = 0$ while $\phi^2_{S,j}(z_S^{(1)})$ is a function of X_1 alone.

Haar distribution of $z_S^{(1)}$. Conditional on $g_{i_1}^{(1)} = h$, each $g_{i_j}^{(1)} \sim \mu$ is independent, so $h^{-1}g_{i_j}^{(1)} \sim \mu$ by left-invariance of Haar. Thus $z_S^{(1)} \sim \mu^{\otimes(k-1)}$ conditionally, and therefore unconditionally. Consequently,

$$\mathbb{E}[\phi^2_{S,j}(z_S^{(1)})] \;=\; \int_{\mathrm{SO}(3)^{k-1}} \phi^2_{S,j}\, d\mu^{\otimes(k-1)} \;=\; \|\phi_{S,j}\|^2_{L^2(\mathrm{SO}(3)^{k-1})} \;=\; 1,$$

since $\{\phi_{S,j}\}$ is an L^2-ONB.

Bounding the f^{*2} term. Pointwise $f^*(X_1)^2 \le \|f^*\|^2_\infty$, so

$$\mathbb{E}[f^*(X_1)^2\phi^2_{S,j}(z_S^{(1)})] \;\le\; \|f^*\|^2_\infty \cdot \mathbb{E}[\phi^2_{S,j}(z_S^{(1)})] \;=\; \|f^*\|^2_\infty.$$

Combining with Lemma 22. Under $s > 3(K-1)/2$, Lemma 22 gives $\|f^*\|_\infty \le C_{\mathrm{sup}}R$. Therefore

$$\mathbb{E}[Y_1^2\phi^2_{S,j}(z_S^{(1)})] \;\le\; \|f^*\|^2_\infty + \sigma^2 \;\le\; C^2_{\mathrm{sup}}R^2 + \sigma^2,$$

and the variance bound follows after dividing by N. □

This is the upper-bound half of Theorem 1 for $\mathcal{F}_{K,\mathrm{inv}}$; the matching lower bound is proved in Appendix G.

Theorem 26 (Upper bound for $\mathcal{F}_{K,\mathrm{inv}}$)**.** *Assume $K \ge 2$ and $s > 3(K-1)/2$. Consider the estimator of Definition 23 using the level-wise common cutoffs*

$$\Lambda_k \;\asymp\; \max\left\{\Lambda_{\min,k},\; \left(\frac{N}{\binom{n}{k}}\right)^{2/(2s+3(k-1))}\right\} \qquad \text{for each } k \in \{2,\dots,K\},$$

where $\Lambda_{\min,k}$ is the threshold constant of Lemma 19 (a positive constant depending on k only, at which $\mathcal{U}^{\mathrm{pure}}_{k,\Lambda}$ is nonempty and $m^{\mathrm{pure}}(\Lambda) \ge c_0\Lambda^{3(k-1)/2}$ holds). Then there exist constants $C > 0$ and $N_0 < \infty$, both depending on n, K, s, R, σ through embedding and mode-count constants (C_1, C_{sup}) but not on $f^* \in \mathcal{F}_{K,\mathrm{inv}}$, such that for all $N \ge N_0$,

$$\mathcal{R}_N(\mathcal{F}_{K,\mathrm{inv}}) \;\le\; C\left(\frac{\binom{n}{K}}{N}\right)^{2s/(2s+3(K-1))}.$$

Equivalently, the displayed rate holds asymptotically as $N \to \infty$ with n, K fixed. The displayed factor $\binom{n}{K}$ isolates the leading subset-count dependence in the rate; constants are not claimed to be uniform in n.

Proof. Write $D_k := 3(k-1)$ and $M_k := \binom{n}{k}$. For each $|S| = k \ge 2$, let $\varphi_S^* := \varphi_S^{\mathrm{pure}}$ denote the pure relative-coordinate representative of f_S^* (Corollary 21).

Step 1: Pathwise orthogonal risk decomposition. By construction, each $T_S\phi_{S,j} \in \mathcal{V}_S$, so $\widehat{f}_S \in \mathcal{V}_S$ a.s. Since $f_S^* \in \mathcal{V}_S$, we have $(\widehat{f}_S - f_S^*) \in \mathcal{V}_S$ a.s. Similarly, $\widehat{f}_0 - f_0^* \in \mathcal{V}_\emptyset$ (constants) a.s. By invariant ANOVA orthogonality (Proposition 11), $\mathcal{V}_\emptyset$ and the $\mathcal{V}_S$ for different S are mutually L^2-orthogonal, so:

$$\|\widehat{f}_N - f^*\|^2_{L^2(\mathbf{G})} \;=\; \|\widehat{f}_0 - f_0^*\|^2_{L^2} + \sum_{k=2}^{K}\sum_{|S|=k} \|\widehat{f}_S - f_S^*\|^2_{L^2} \qquad \text{a.s.}$$

Taking expectations:

$$\mathbb{E}\|\widehat{f}_N - f^*\|^2_{L^2} \;=\; \mathbb{E}\|\widehat{f}_0 - f_0^*\|^2 \;+\; \sum_{k=2}^{K}\sum_{|S|=k} \mathbb{E}\|\widehat{f}_S - f_S^*\|^2_{L^2}.$$

Step 2: Constant-term risk. $\widehat{f}_0 = \bar{Y}$ is unbiased for $f_0^* = \int f^*\, d\mu^{\otimes n}$, with

$$\mathbb{E}\|\widehat{f}_0 - f_0^*\|^2 \;=\; \mathrm{Var}(\bar{Y}) \;=\; \frac{\mathrm{Var}(f^*(X_1)) + \sigma^2}{N} \;\le\; \frac{R^2 + \sigma^2}{N},$$

using $\mathrm{Var}(f^*(X_1)) \le \|f^*\|_{L^2}^2 \le \|f^*\|_{H^s}^2 \le R^2$.

Step 3: Per-component bias-variance decomposition. Fix $|S| = k \in \{2, \dots, K\}$. The mean of the component estimator equals the pulled-back spectral projection of φ_S^*:

$$\mathbb{E}[\widehat{f}_S] \;=\; \sum_{j=1}^{M_S} \mathbb{E}[\widehat{a}_{S,j}] \cdot T_S\phi_{S,j} \;\overset{\text{Prop. 24}}{=}\; \sum_{j=1}^{M_S} \langle \varphi_S^*, \phi_{S,j}\rangle \cdot T_S\phi_{S,j} \;=\; T_S(\Pi_{\Lambda_k}^{\mathrm{rel}}\varphi_S^*),$$

where $\Pi_{\Lambda_k}^{\mathrm{rel}}$ is the $L^2(\mathrm{SO}(3)^{k-1})$-orthogonal projection onto the retained full-support $P_\Delta^\perp$-range modes at relative Casimir level $\le \Lambda_k$.

Bias (on the relative domain). Since $f_S^* = T_S\varphi_S^*$ (Corollary 21),

$$\big\|\mathbb{E}[\widehat{f}_S] - f_S^*\big\|_{L^2(\mathbf{G})}^2 \;=\; \big\|T_S(\Pi_{\Lambda_k}^{\mathrm{rel}}\varphi_S^* - \varphi_S^*)\big\|_{L^2(\mathbf{G})}^2 \;\overset{\text{Lem. 14}}{=}\; \big\|\Pi_{\Lambda_k}^{\mathrm{rel}}\varphi_S^* - \varphi_S^*\big\|_{L^2(\mathrm{SO}(3)^{k-1})}^2.$$

Why Lemma 7 applies to $\Pi_{\Lambda_k}^{\mathrm{rel}}$. Note that $\Pi_{\Lambda_k}^{\mathrm{rel}}$ is a *restricted* projector, onto retained *full-support* $P_\Delta^\perp$*-range* modes at relative Casimir level $\le \Lambda_k$ rather than onto all modes at that Casimir level. However, by Corollary 21, φ_S^* already lies in the full-support $P_\Delta^\perp$-range subspace, and this subspace is Peter–Weyl blockwise invariant (full-support is a mode-by-mode condition; $P_\Delta^\perp$ is block-diagonal by Lemma 16). Therefore, within this invariant subspace, $\Pi_{\Lambda_k}^{\mathrm{rel}}$ coincides with ordinary spectral truncation at Casimir $\le \Lambda_k$ applied to φ_S^*. Consequently, the standard tail bound of Lemma 7 applies directly:

$$\big\|\Pi_{\Lambda_k}^{\mathrm{rel}}\varphi_S^* - \varphi_S^*\big\|_{L^2(\mathrm{SO}(3)^{k-1})}^2 \;\le\; (1+\Lambda_k)^{-s}\,\|\varphi_S^*\|_{H^s(\mathrm{SO}(3)^{k-1})}^2.$$

This step is cleaner than an ambient-$\mathbf{G}$-Casimir truncation because the estimator's basis is defined by the relative-Casimir cutoff $\le \Lambda_k$ on $\mathrm{SO}(3)^{k-1}$; $\Pi_{\Lambda_k}^{\mathrm{rel}}$ is exactly the projection the estimator implements.

Variance. Since $\{T_S\phi_{S,j}\}_{j=1}^{M_S}$ is $L^2(\mathbf{G})$-orthonormal (Definition 23):

$$\mathbb{E}\big\|\widehat{f}_S - \mathbb{E}[\widehat{f}_S]\big\|_{L^2(\mathbf{G})}^2 \;=\; \sum_{j=1}^{M_S} \mathrm{Var}(\widehat{a}_{S,j}) \;\le\; M_S \cdot \frac{C_{\sup}^2 R^2 + \sigma^2}{N} \;\le\; C_1\Lambda_k^{D_k/2} \cdot \frac{C_{\sup}^2 R^2 + \sigma^2}{N},$$

using Proposition 25 and the mode count $M_S \le C_1\Lambda_k^{D_k/2}$ from Lemma 19.

Combining (bias-variance decomposition).

$$\begin{aligned}\mathbb{E}\|\widehat{f}_S - f_S^*\|_{L^2}^2 \;&=\; \|\mathbb{E}[\widehat{f}_S] - f_S^*\|_{L^2}^2 + \mathbb{E}\|\widehat{f}_S - \mathbb{E}[\widehat{f}_S]\|_{L^2}^2 \\ &\le\; (1+\Lambda_k)^{-s}\|\varphi_S^*\|_{H^s}^2 + C_1\Lambda_k^{D_k/2}\frac{C_{\sup}^2 R^2 + \sigma^2}{N}.\end{aligned}$$

Step 4: Sum over subsets at level k using the global budget. Summing over $|S| = k$ and applying the global budget from Corollary 21 ($\sum_{k=2}^K \sum_{|S|=k} \|\varphi_S^*\|_{H^s}^2 \le R^2$):

$$\begin{aligned}\sum_{|S|=k} \mathbb{E}\|\widehat{f}_S - f_S^*\|_{L^2}^2 \;&\le\; (1+\Lambda_k)^{-s} \sum_{|S|=k} \|\varphi_S^*\|_{H^s}^2 + M_k \cdot C_1\Lambda_k^{D_k/2}\frac{C_{\sup}^2 R^2 + \sigma^2}{N} \\ &\le\; (1+\Lambda_k)^{-s}R^2 + C_1 M_k \Lambda_k^{D_k/2}\frac{C_{\sup}^2 R^2 + \sigma^2}{N}.\end{aligned}$$

Step 5: Common cutoff balances bias and variance. Choose Λ_k to balance the two terms above: $(1+\Lambda_k)^{-s}R^2 \asymp M_k\Lambda_k^{D_k/2}(C_{\sup}^2 R^2 + \sigma^2)/N$. Solving:

$$\Lambda_k \;\asymp\; \left(\frac{NR^2}{M_k(C_{\sup}^2 R^2 + \sigma^2)}\right)^{2/(2s+D_k)} \;\asymp\; \left(\frac{N}{M_k}\right)^{2/(2s+D_k)},$$

with the $R, \sigma, C_{\sup}$ factors absorbed into the implicit constant. For small N (specifically $N < M_k$), the displayed expression may fall below the smallest Casimir level $\Lambda_{\min,k}$ at which the retained basis subspace is nonempty; we then fix $\Lambda_k := \Lambda_{\min,k}$ (a positive constant independent of N). This only affects constants in the non-asymptotic regime. In the asymptotic regime $N \to \infty$ with n, K fixed, the displayed scaling dominates. This cutoff depends only on N, n, k, s, and the known universal constants $R, \sigma, C_{\sup}, C_1$—not on any target-dependent quantities such as $\|\varphi_S^*\|_{H^s}$. Substituting back:

$$\sum_{|S|=k} \mathbb{E}\|\widehat{f}_S - f_S^*\|_{L^2}^2 \;\le\; C\left(\frac{M_k}{N}\right)^{2s/(2s+D_k)},$$

where $C = C(s, K, n, R, \sigma, C_1, C_{\sup})$ is independent of f^*.

Step 6: Top-order dominance. Summing across levels $k = 2, \dots, K$ and adding the constant term from Step 2:

$$\mathbb{E}\|\widehat{f}_N - f^*\|_{L^2}^2 \;\le\; \frac{R^2+\sigma^2}{N} + C\sum_{k=2}^{K}\left(\frac{M_k}{N}\right)^{2s/(2s+D_k)}.$$

For $k < K$: $D_k = 3(k-1) < 3(K-1) = D_K$, hence $2s/(2s+D_k) > 2s/(2s+D_K)$. With n and K fixed, $M_k = \binom{n}{k}$ is a constant, so

$$\left(\frac{M_k}{N}\right)^{2s/(2s+D_k)} \;=\; o\left(\left(\frac{M_K}{N}\right)^{2s/(2s+D_K)}\right) \quad \text{as } N \to \infty.$$

The constant term $O(N^{-1})$ also decays faster than $N^{-2s/(2s+D_K)}$ (since $2s/(2s+D_K) < 1$ when $D_K > 0$, i.e., when $K \ge 2$). Therefore the top body-order level $k = K$ dominates:

$$\mathbb{E}\|\widehat{f}_N - f^*\|_{L^2}^2 \;\le\; C'\left(\frac{\binom{n}{K}}{N}\right)^{2s/(2s+3(K-1))}.$$

□

G Lower Bound for $\mathcal{F}_{K,\mathrm{inv}}$

The lower bound follows the information-theoretic reduction to multiple testing [37, 32]: construct a finite packing of $\mathcal{F}_{K,\mathrm{inv}}$ whose elements are jointly hard to distinguish from noise, then apply Fano's inequality. The technical novelty lies in the packing construction, which requires Peter–Weyl analysis on the pure-S invariant ANOVA subspaces (Theorem 18) and a mode-count lower bound (Lemma 19).

Lemma 27 (Fano; Tsybakov Thm 2.5)**.** *Let $\{\theta_0, \theta_1, \dots, \theta_L\}$ be hypotheses in a parameter space with semi-metric d, and let $\{P_{\theta_a}\}_{a=0}^{L}$ be the corresponding distributions. If $d(\theta_a, \theta_b) \ge 2\delta$ for all $0 \le a < b \le L$ and $\frac{1}{L}\sum_{a=1}^{L} \mathrm{KL}(P_{\theta_a} \| P_{\theta_0}) \le \beta$ with $\log L \ge 2\beta + \log 2$, then*

$$\inf_{\widehat{\theta}} \max_{0 \le a \le L} P_{\theta_a}\big(d(\widehat{\theta}, \theta_a) \ge \delta\big) \;\ge\; \tfrac{1}{4}.$$

Lemma 28 (Varshamov–Gilbert; Tsybakov Lem 2.9)**.** *For p sufficiently large, there exists $\Omega = \{\omega^{(0)}, \omega^{(1)}, \dots, \omega^{(L)}\} \subset \{0,1\}^p$ with $\omega^{(0)} = \mathbf{0}$ and $L \ge 2^{p/8}$ such that*

$$d_H(\omega^{(a)}, \omega^{(b)}) \;\ge\; p/8 \qquad \textit{for all } 0 \le a < b \le L.$$

In particular, every nonzero codeword $\omega \in \Omega \setminus \{\mathbf{0}\}$ satisfies $|\omega|_0 \ge p/8$.

Lemma 29 (Packing orthonormal system)**.** *Let $\{S_j\}_{j=1}^{M}$ ($M = \binom{n}{K}$) enumerate the K-subsets of $[n]$, and let $m \le m^{\mathrm{pure}}(\Lambda)$. Choose any L^2-orthonormal family $\{\psi_l\}_{l=1}^{m} \subset \mathcal{U}_{K,\Lambda}^{\mathrm{pure}}$ (a subset of size m from a complete ONB of the full-support $P_\Delta^\perp$-range subspace at Casimir level $\le \Lambda$; nonemptiness requires $\Lambda \ge \Lambda_{\min,K}$, as established in Lemma 19). For each j, l define*

$$\Psi_{j,l} \;:=\; T_{S_j}^{\mathrm{pure}}\psi_l \;=\; T_{S_j}\psi_l,$$

where the second equality holds because $\psi_l \in \mathrm{range}(P_\Delta^\perp)$ and $T_{S_j}^{\mathrm{pure}} = T_{S_j} \circ P_\Delta^\perp$ (Theorem 18). Then $\Psi_{j,l} \in \mathcal{V}_{S_j}$ and $\{\Psi_{j,l}\}_{1 \le j \le M,\, 1 \le l \le m}$ is an $L^2(\mathbf{G})$-orthonormal family of size mM.

Proof. $\Psi_{j,l} \in \mathcal{V}_{S_j}$ by Theorem 18 (full-support $P_\Delta^\perp$-range pulls back into $\mathcal{V}_{S_j}$).

For $j \neq j'$: by invariant ANOVA orthogonality (Proposition 11), $\mathcal{V}_{S_j} \perp \mathcal{V}_{S_{j'}}$ in $L^2(\mathbf{G})$, so $\langle \Psi_{j,l}, \Psi_{j',l'} \rangle_{L^2(\mathbf{G})} = 0$.

For $j = j'$: since $\psi_l \in \mathrm{range}(P_\Delta^\perp)$, the identity $\Psi_{j,l} = T_{S_j}\psi_l$ allows Lemma 14 (T_{S_j} is an L^2-isometry) to apply directly:

$$\langle \Psi_{j,l}, \Psi_{j,l'} \rangle_{L^2(\mathbf{G})} \;=\; \langle T_{S_j}\psi_l, T_{S_j}\psi_{l'} \rangle_{L^2(\mathbf{G})} \;=\; \langle \psi_l, \psi_{l'} \rangle_{L^2(\mathrm{SO}(3)^{K-1})} \;=\; \delta_{l,l'}.$$

In particular, $\|\Psi_{j,l}\|_{L^2}^2 = 1$. □

This is the lower-bound half of Theorem 1 for $\mathcal{F}_{K,\mathrm{inv}}$; combined with the upper bound of Appendix F, it establishes the minimax rate.

Theorem 30 (Lower bound for $\mathcal{F}_{K,\mathrm{inv}}$)**.** *Assume $K \geq 2$ and $s > 3(K-1)/2$, and consider the Gaussian regression model with noise variance σ^2 and design $X \sim \mu^{\otimes n}$. Then there exist constants $c > 0$ and $N_0 < \infty$, depending on n, K, s, R, σ only (independent of N and of any specific target $f^* \in \mathcal{F}_{K,\mathrm{inv}}$), such that for all $N \geq N_0$,*

$$\mathcal{R}_N(\mathcal{F}_{K,\mathrm{inv}}) \;\geq\; c\left(\frac{\binom{n}{K}}{N}\right)^{2s/(2s+3(K-1))}.$$

Equivalently, the displayed rate holds asymptotically as $N \to \infty$ with n, K fixed.

Proof. Let $D := 3(K-1)$, $M := \binom{n}{K}$, and $m := \lfloor c_0 \Lambda^{D/2}/2 \rfloor$, where c_0 is the constant in the mode-count lower bound $m^{\mathrm{pure}}(\Lambda) \geq c_0\Lambda^{D/2}$ (Lemma 19). Set $p := mM$, so $m \asymp \Lambda^{D/2}$ and $p \asymp M\Lambda^{D/2}$.

Step 1: Sobolev control on the packing subspace. Let $\{\psi_l\}_{l=1}^m$ be an L^2-orthonormal family in $\mathcal{U}_{K,\Lambda}^{\mathrm{pure}}$ and $\{\Psi_{j,l}\}$ the pulled-back family of Lemma 29. Consider any linear combination

$$g \;=\; \sum_{j=1}^{M}\sum_{l=1}^{m} a_{j,l}\,\Psi_{j,l}, \qquad \text{and set } \phi_j := \sum_{l=1}^{m} a_{j,l}\,\psi_l \in \mathcal{U}_{K,\Lambda}^{\mathrm{pure}}.$$

Since each $\phi_j \in \mathrm{range}(P_\Delta^\perp)$, we have $T_{S_j}^{\mathrm{pure}}\phi_j = T_{S_j}\phi_j$, so $g = \sum_j T_{S_j}\phi_j$. Different S_j's produce functions with mutually disjoint Peter–Weyl supports in $L^2(\mathbf{G})$ ($T_{S_j}\phi_j$ has Fourier support in blocks $\boldsymbol{\ell}$ with $\ell_i \geq 1$ exactly for $i \in S_j$, and these conditions select disjoint sets of blocks for distinct S_j's), so the $H^s(\mathbf{G})$-norms decompose blockwise:

$$\|g\|_{H^s(\mathbf{G})}^2 \;=\; \sum_{j=1}^{M}\|T_{S_j}\phi_j\|_{H^s(\mathbf{G})}^2 \;\leq\; K^s\sum_{j=1}^{M}\|\phi_j\|_{H^s(\mathrm{SO}(3)^{K-1})}^2,$$

where the inequality is Theorem 20 with $k = K$. Since each ϕ_j is supported at Casimir level $\leq \Lambda$, $\|\phi_j\|_{H^s}^2 \leq (1+\Lambda)^s\|\phi_j\|_{L^2}^2$. Using $\|\phi_j\|_{L^2}^2 = \sum_l a_{j,l}^2$ (since $\{\psi_l\}$ is L^2-orthonormal) and $\|g\|_{L^2}^2 = \sum_{j,l} a_{j,l}^2$ (Lemma 29):

$$\|g\|_{H^s(\mathbf{G})}^2 \;\leq\; K^s(1+\Lambda)^s\,\|g\|_{L^2(\mathbf{G})}^2. \tag{4}$$

Step 2: Hypothesis family. By Lemma 28, there exists $\Omega = \{\omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(L)}\} \subset \{0,1\}^p$ with $\omega^{(0)} = \mathbf{0}$, $L \geq 2^{p/8}$, and pairwise Hamming distance $\geq p/8$. For $\omega \in \Omega$, define

$$f_\omega \;:=\; \eta\sum_{j=1}^{M}\sum_{l=1}^{m}\omega_{(j-1)m+l}\,\Psi_{j,l}.$$

In particular, $f_{\mathbf{0}} \equiv 0$ is one of the hypotheses. Write $P_{\mathbf{0}} := P_{f_{\mathbf{0}}}$ (the N-sample distribution under the zero function) and $P_\omega := P_{f_\omega}$.

Step 3: Choice of Λ and η. Set

$$\Lambda \;:=\; c_\Lambda \left(\frac{NR^2}{M\sigma^2}\right)^{2/(D+2s)}, \qquad \eta^2 \;:=\; c_\eta \,\frac{\sigma^2}{N},$$

where $c_\Lambda > 0$ is any fixed positive constant, and $c_\eta > 0$ is chosen sufficiently small so that the membership constraint (Step 4) and the Fano KL budget (Step 7) both hold.

Step 4: Membership ($f_\omega \in \mathcal{F}_{K,\mathrm{inv}}$). By Lemma 29, $\|f_\omega\|_{L^2}^2 = \eta^2|\omega|_0 \le \eta^2 p$. Applying (4) with $g = f_\omega$:

$$\|f_\omega\|_{H^s(\mathbf{G})}^2 \;\le\; K^s(1+\Lambda)^s\,\eta^2 p \;\lesssim\; c_\eta K^s\,\frac{\sigma^2}{N}\,M\Lambda^{D/2}(1+\Lambda)^s \;\asymp\; c_\eta\, c_\Lambda^{s+D/2}\,K^s R^2,$$

where the last $\asymp$ uses the definition of Λ: $\Lambda^{D/2+s} = c_\Lambda^{D/2+s}\cdot(NR^2/(M\sigma^2))$, so $M\Lambda^{D/2}(1+\Lambda)^s \asymp M\Lambda^{D/2+s} \asymp c_\Lambda^{D/2+s}\cdot NR^2/\sigma^2$, and $(1+\Lambda)^s \asymp \Lambda^s$ for $\Lambda \ge 1$. Choosing $c_\eta > 0$ sufficiently small (for the fixed choice of c_Λ and given K, R, σ) yields $\|f_\omega\|_{H^s}^2 \le R^2$. Moreover, each $\Psi_{j,l} \in \mathcal{V}_{S_j}$ with $|S_j| = K$, so f_ω lies in the degree-K part of the invariant ANOVA decomposition; combined with the H^s-bound, $f_\omega \in \mathcal{F}_{K,\mathrm{inv}}$.

Step 5: Separation. By Lemma 29's L^2-orthonormality of $\{\Psi_{j,l}\}$:

$$\|f_\omega - f_{\omega'}\|_{L^2(\mathbf{G})}^2 \;=\; \eta^2\, d_H(\omega,\omega') \;\ge\; \eta^2\, p/8 \qquad \text{for all } \omega \ne \omega' \in \Omega.$$

This holds in particular for $\omega' = \mathbf{0}$, i.e., $\|f_\omega - f_{\mathbf{0}}\|_{L^2}^2 = \eta^2|\omega|_0 \ge \eta^2 p/8$ for every $\omega \ne \mathbf{0}$, so the Fano separation radius $2\delta := \eta\sqrt{p/8}$ holds uniformly among *all* pairs in Ω—including the null reference.

Step 6: KL under the Gaussian regression model. For the Gaussian observation model $Y_\alpha = f^*(X_\alpha) + \xi_\alpha$ with $\xi_\alpha \sim \mathcal{N}(0,\sigma^2)$ i.i.d. and $X_\alpha \sim \mu^{\otimes n}$, the N-sample KL divergence is

$$\mathrm{KL}(P_\omega \,\|\, P_{\mathbf{0}}) \;=\; \frac{N}{2\sigma^2}\,\|f_\omega\|_{L^2(\mathbf{G})}^2 \;=\; \frac{N\eta^2\,|\omega|_0}{2\sigma^2} \;\le\; \frac{N\eta^2 p}{2\sigma^2} \;=\; \frac{c_\eta p}{2}.$$

Hence the averaged KL against the null reference $\theta_0 = f_{\mathbf{0}}$ satisfies

$$\bar\beta \;:=\; \frac{1}{L}\sum_{a=1}^{L}\mathrm{KL}(P_{\omega^{(a)}} \,\|\, P_{\mathbf{0}}) \;\le\; \frac{c_\eta p}{2}.$$

Step 7: Fano condition. Lemma 27 requires $\log L \ge 2\bar\beta + \log 2$. We have $\log L \ge (p/8)\log 2$ (VG) and $\bar\beta \le c_\eta p/2$ (Step 6). Choosing $c_\eta \le (\log 2)/32$ gives $\bar\beta \le p(\log 2)/64$, hence

$$(p/8)\log 2 \;\ge\; 2\cdot p(\log 2)/64 + \log 2 \;=\; p(\log 2)/32 + \log 2,$$

i.e., $(3p/32)\log 2 \ge \log 2$, which holds for $p \ge 32/3$ (so $p \ge 11$). Under Step 3's choice of Λ, $p = mM \asymp M\Lambda^{D/2} \to \infty$ as $N \to \infty$ with n, K fixed, so $p \ge 11$ holds for all $N \ge N_0(n, K, s, R, \sigma)$.

Step 8: From Fano probability to expected risk. Lemma 27 yields

$$\inf_{\widehat f}\max_{\omega\in\Omega} P_\omega\big(\|\widehat f - f_\omega\|_{L^2} \ge \delta\big) \;\ge\; \tfrac14.$$

For any nonnegative random variable Z, $\mathbb{E}[Z^2] \ge \delta^2\,\Pr(Z \ge \delta)$ (applying Markov's inequality to Z^2 at threshold δ^2). Applying this with $Z = \|\widehat f - f_\omega\|_{L^2}$ and taking max over $\omega \in \Omega$ and inf over $\widehat f$:

$$\inf_{\widehat f}\max_{\omega\in\Omega}\mathbb{E}_\omega\|\widehat f - f_\omega\|_{L^2}^2 \;\ge\; \frac{\delta^2}{4}.$$

Since all $f_\omega \in \mathcal{F}_{K,\mathrm{inv}}$ (Step 4), the minimax risk satisfies $\mathcal{R}_N(\mathcal{F}_{K,\mathrm{inv}}) \ge \delta^2/4 = \eta^2 p/128$.

Step 9: Rate. Substituting $\eta^2 = c_\eta\sigma^2/N$ and $p \asymp M\Lambda^{D/2}$ with $\Lambda \asymp (NR^2/(M\sigma^2))^{2/(D+2s)}$ into Step 8:

$$\mathcal{R}_N(\mathcal{F}_{K,\mathrm{inv}}) \;\ge\; \frac{\delta^2}{4} \;\asymp\; \eta^2 p \;\asymp\; \frac{\sigma^2}{N}\cdot M\Lambda^{D/2} \;\asymp\; \sigma^2\left(\frac{M}{N}\right)^{2s/(2s+D)}.$$

Absorbing $\sigma^2, R, K, c_0, c_\Lambda, c_\eta$ into a single constant c depending on n, K, s, R, σ:

$$\mathcal{R}_N(\mathcal{F}_{K,\mathrm{inv}}) \;\geq\; c\left(\frac{\binom{n}{K}}{N}\right)^{2s/(2s+3(K-1))} \qquad \text{for all } N \geq N_0,$$

which matches the theorem statement. □

Remark 31 (Estimator-independent lower bounds and KRR)**.** *The Fano argument in Step 8 takes the infimum over all measurable estimators, so the minimax lower bound applies to KRR as well as to spectral projection; Theorem 26 establishes the matching upper bound using spectral projection. Our analysis also provides geometric ingredients for studying KRR: Corollary 21 constructs pure relative-coordinate representatives, Theorem 20 controls Sobolev norms under pullback, and Lemma 19 establishes the pure-mode count. Obtaining a matching KRR upper bound would additionally require a suitable kernel, an RKHS–Sobolev relationship, covariance-operator eigenvalue decay, a regularization schedule, and enforcement of the $P_\Delta^\perp$ constraint. Tahmasebi and Jegelka [30] provide related KRR theory for invariance alone; for the diagonal* $\mathrm{SO}(3)$ *action, their orbit-dimension principle gives the same three-dimensional reduction as Proposition 32. Extending KRR to body-order ANOVA with pure-S constraints remains an open direction.*

H Parallel Rates

This section derives the minimax rates for the three remaining function classes by adapting the templates of Appendices F and G to reduced or alternate effective dimensions.

Convention for this section. All rate statements below are asymptotic for fixed n and K as $N \to \infty$. Implicit constants may depend on n, K, s, R, σ through embedding and mode-count constants, but not on N or on the target function f^*. The same real orthonormal basis convention as Appendix F is used throughout. For each cutoff Λ, $\Lambda_{\min}$ denotes the corresponding threshold constant (as in Lemma 19, generalized to the appropriate subspace and domain) at which the relevant basis subspace is nonempty; $\Lambda_{\min}$ is a positive constant independent of N.

Proposition 32 ($\mathcal{F}_{\mathrm{inv}}$)**.** $\mathcal{R}_N(\mathcal{F}_{\mathrm{inv}}) \asymp N^{-2s/(2s+3(n-1))}$ *for $s > 3(n-1)/2$, as $N \to \infty$ with n fixed.*

Proof. Every invariant function on $\mathrm{SO}(3)^n$ admits a representation $f(\mathbf{g}) = \varphi(g_1^{-1}g_2, \ldots, g_1^{-1}g_n)$ for some φ on $\mathrm{SO}(3)^{n-1}$, with $\|f\|_{H^s(\mathbf{G})}$ and $\|\varphi\|_{H^s(\mathrm{SO}(3)^{n-1})}$ bilaterally bounded by constants depending only on n (Sobolev equivalence, Theorem 20 applied at $k = n$, $S = [n]$). Effective dimension is $3(n-1)$; there is a single subspace, so $M = 1$. No pure-S projection is needed here because $\mathcal{F}_{\mathrm{inv}}$ is not decomposed by body-order; it is treated as a single invariant function on the quotient $\mathrm{SO}(3)^n/\mathrm{SO}(3) \cong \mathrm{SO}(3)^{n-1}$ (contrasted with $\mathcal{F}_{K,\mathrm{inv}}$, where the body-order decomposition requires the pure-representative corollary). Throughout, we use the same real orthonormal basis convention as Appendix F.

Upper bound. Apply the spectral projection estimator on $\mathrm{SO}(3)^{n-1}$ with Casimir cutoff Λ; the mode count is $\leq C\Lambda^{3(n-1)/2}$ (Lemma 5). Bias $\lesssim (1+\Lambda)^{-s}\|\varphi\|_{H^s}^2$, variance $\lesssim \Lambda^{3(n-1)/2}(\|f\|_\infty^2+\sigma^2)/N$ by the same coefficient-wise variance calculation as Proposition 25, now applied to an L^2-orthonormal spectral basis on $\mathrm{SO}(3)^{n-1}$, with $\|f\|_\infty = \|\varphi\|_\infty \leq C_{\mathrm{emb}}(s, n-1)\cdot R$ from Lemma 6 (applied to φ on $\mathrm{SO}(3)^{n-1}$) and Theorem 20 (lower bound $\|\varphi\|_{H^s} \leq \|f\|_{H^s}$), under $s > 3(n-1)/2$. Choose $\Lambda \asymp \max\{\Lambda_{\min}, N^{2/(2s+3(n-1))}\}$; in the asymptotic regime this gives rate $N^{-2s/(2s+3(n-1))}$.

Lower bound. Apply the Varshamov–Gilbert–Fano template of Theorem 30 with $M = 1$ subspace and $m \asymp \Lambda^{3(n-1)/2}$ Peter–Weyl modes on $\mathrm{SO}(3)^{n-1}$ (via Lemma 5). For ψ supported at Casimir level $\leq \Lambda$, Theorem 20 with $k = n$, $S = [n]$ and the spectral bound $\|\psi\|_{H^s}^2 \leq (1+\Lambda)^s\|\psi\|_{L^2}^2$ yield the $H^s(\mathbf{G})$ membership bound

$$\|T_{[n]}\psi\|_{H^s(\mathbf{G})}^2 \;\leq\; n^s(1+\Lambda)^s\|\psi\|_{L^2}^2.$$

Balancing KL and membership constraints as in Theorem 30 gives $\delta^2 \asymp N^{-2s/(2s+3(n-1))}$. □

Proposition 33 ($\mathcal{F}_0$)**.** $\mathcal{R}_N(\mathcal{F}_0) \asymp N^{-2s/(2s+3n)}$ *for $s > 3n/2$, as $N \to \infty$ with n fixed.*

Proof. No structural reduction; the domain is $\mathrm{SO}(3)^n$ with dimension $3n$. We use the same real orthonormal basis convention as Appendix F.

Upper bound. Apply the spectral projection estimator directly on $\mathrm{SO}(3)^n$ with Casimir cutoff Λ; the mode count is $M_\Lambda \le C\Lambda^{3n/2}$ (Lemma 5). Bias $\lesssim (1+\Lambda)^{-s}\|f^*\|_{H^s}^2$, variance $\lesssim M_\Lambda(\|f^*\|_\infty^2 + \sigma^2)/N$ by the same coefficient-wise variance calculation as Proposition 25, now applied directly to an L^2-orthonormal spectral basis on $\mathrm{SO}(3)^n$, with $\|f^*\|_\infty \le C_{\mathrm{emb}}(s,n)\cdot R$ from Lemma 6 under $s > 3n/2$. Choose $\Lambda \asymp \max\{\Lambda_{\min}, N^{2/(2s+3n)}\}$; in the asymptotic regime this gives rate $N^{-2s/(2s+3n)}$.

Lower bound. Apply the Varshamov–Gilbert–Fano template of Theorem 30 with $M = 1$ subspace (no subset combinatorics) and $m \asymp \Lambda^{3n/2}$ Peter–Weyl modes on $\mathrm{SO}(3)^n$. Balancing membership ($\eta^2 m(1+\Lambda)^s \le R^2$) and KL ($\eta^2 \le c\sigma^2/N$) constraints yields $\Lambda \asymp (NR^2/\sigma^2)^{2/(2s+3n)}$, giving separation $\delta^2 \asymp N^{-2s/(2s+3n)}$. □

Lemma 34 (Sup-norm bound for $\mathcal{F}_K$)**.** *Assume $K \ge 1$ and $s > 3K/2$. If $f \in \mathcal{F}_K$ and $\|f\|_{H^s(\mathbf{G})} \le R$, then*

$$\|f\|_\infty \;\le\; C_{\sup,K}\, R, \qquad C_{\sup,K} \;:=\; \sqrt{\sum_{k=0}^{K}\binom{n}{k}}\cdot C_{\mathrm{emb}}(s,K).$$

Proof. Decompose $f = f_0 + \sum_{k=1}^K \sum_{|S|=k} f_S$ with $f_S \in \mathcal{W}_S$ (all k-body components from $k=1$ to K contribute in the non-invariant class $\mathcal{F}_K$).

Constant component ($k=0$). Since f_0 is constant on $\mathbf{G}$,

$$\|f_0\|_\infty \;=\; \|f_0\|_{L^2} \;=\; \|f_0\|_{H^s} \;\le\; C_{\mathrm{emb}}(s,K)\,\|f_0\|_{H^s},$$

using $C_{\mathrm{emb}}(s,K) \ge 1$ (the $\boldsymbol{\ell} = \mathbf{0}$ term contributes 1 to the defining series and all other terms are nonnegative).

k-body components with $|S| = k \ge 1$. $f_S \in \mathcal{W}_S$ depends only on coordinates in S and thus can be viewed as a function on $\mathrm{SO}(3)^k$ (dimension $3k \le 3K$). By Lemma 6 with $m = k \le K$ and $s > 3K/2 \ge 3m/2$:

$$\|f_S\|_\infty \;\le\; C_{\mathrm{emb}}(s,k)\,\|f_S\|_{H^s} \;\le\; C_{\mathrm{emb}}(s,K)\,\|f_S\|_{H^s},$$

since $C_{\mathrm{emb}}(s,m)$ is non-decreasing in m (the defining series on $\mathbb{Z}_{\ge 0}^m$ is embedded in the one on $\mathbb{Z}_{\ge 0}^K$ via trivial modes).

Combining across subsets. Let $M := \sum_{k=0}^K \binom{n}{k}$. By the triangle inequality and Cauchy–Schwarz:

$$\|f\|_\infty \;\le\; \sum_{|S|\le K}\|f_S\|_\infty \;\le\; C_{\mathrm{emb}}(s,K)\sum_{|S|\le K}\|f_S\|_{H^s} \;\le\; C_{\mathrm{emb}}(s,K)\sqrt{M}\sqrt{\sum_{|S|\le K}\|f_S\|_{H^s}^2}.$$

By general ANOVA H^s-orthogonality (Proposition 10), $\sum_{|S|\le K}\|f_S\|_{H^s}^2 \le \|f\|_{H^s}^2 \le R^2$. Therefore $\|f\|_\infty \le C_{\sup,K}\,R$. □

Proposition 35 ($\mathcal{F}_K$)**.** $\mathcal{R}_N(\mathcal{F}_K) \asymp \left(\binom{n}{K}/N\right)^{2s/(2s+3K)}$ *for $s > 3K/2$, as $N \to \infty$ with n, K fixed.*

Proof. Without invariance, the k-body ANOVA component f_S for $|S| = k$ depends on all k original coordinates $(g_{i_1}, \ldots, g_{i_k}) \in \mathrm{SO}(3)^k$—no reduction to relative coordinates occurs. The effective dimension per k-body component is therefore $3k$.

Upper bound. For each $|S| = k \in \{1, \ldots, K\}$ at common cutoff Λ_k (specified below), construct a spectral projection estimator directly on $\mathrm{SO}(3)^k$: the basis $\{\phi_{S,j}^{\mathrm{abs}}\}_{j=1}^{M_S}$ consists of real L^2-orthonormal Peter–Weyl product modes on $\mathrm{SO}(3)^k$ whose Fourier support is exactly S (all k coordinates carry $\ell_i \ge 1$), forming an ONB of $\mathcal{W}_S$ (Definition 9) at spectral level $\le \Lambda_k$, with $M_S \le C\Lambda_k^{3k/2}$ (Lemma 5). Coefficients: $\widehat{a}_{S,j} := \frac{1}{N}\sum_{\alpha=1}^N Y_\alpha \phi_{S,j}^{\mathrm{abs}}(g_S^{(\alpha)})$. Since $\phi_{S,j}^{\mathrm{abs}} \in \mathcal{W}_S$, the component

estimator $\widehat{f}_S := \sum_j \widehat{a}_{S,j}\phi^{\mathrm{abs}}_{S,j} \in \mathcal{W}_S$ a.s. Constant term: $\widehat{f}_0 := \bar{Y}$. By general ANOVA orthogonality (Proposition 10), the $\mathcal{W}_S$ are mutually $L^2(\mathbf{G})$-orthogonal across S, so the risk decomposes pathwise:

$$\|\widehat{f}_N - f^*\|^2_{L^2} \;=\; \|\widehat{f}_0 - f^*_0\|^2 + \sum_{k=1}^{K}\sum_{|S|=k} \|\widehat{f}_S - f^*_S\|^2_{L^2} \quad \text{a.s.}$$

Variance per coefficient: $\mathrm{Var}(\widehat{a}_{S,j}) \le (C^2_{\sup,K}R^2 + \sigma^2)/N$ by the same argument as Proposition 25 (Haar-uniform $g^{(1)}_S$, L^2-normalized $\phi^{\mathrm{abs}}_{S,j}$, sup-norm control from Lemma 34). Bias: since $f^*_S \in \mathcal{W}_S$ and $\mathcal{W}_S$ (full-support subspace of $L^2(\mathrm{SO}(3)^k)$) is Peter–Weyl blockwise invariant, the restricted projector Π_{Λ_k} onto retained $\mathcal{W}_S$-modes at Casimir $\le \Lambda_k$ coincides with ordinary spectral truncation within $\mathcal{W}_S$ applied to f^*_S. Hence Lemma 7 applies directly on $\mathrm{SO}(3)^k$ (no pullback factor, since no relative-coordinate reduction occurs):

$$\|f^*_S - \Pi_{\Lambda_k} f^*_S\|^2_{L^2(\mathbf{G})} \;\le\; (1+\Lambda_k)^{-s}\|f^*_S\|^2_{H^s}.$$

Mode count: $M_S \le C\Lambda_k^{3k/2}$ on $\mathrm{SO}(3)^k$ (Lemma 5).

Level-k risk via global budget. Summing across $|S| = k$ and applying the H^s-budget from Proposition 10 ($\sum_{k=0}^{K}\sum_{|S|=k}\|f^*_S\|^2_{H^s} \le R^2$):

$$\begin{aligned}\sum_{|S|=k}\mathbb{E}\|\widehat{f}_S - f^*_S\|^2_{L^2} \;&\le\; (1+\Lambda_k)^{-s}\sum_{|S|=k}\|f^*_S\|^2_{H^s} + \binom{n}{k} C\Lambda_k^{3k/2}\frac{C^2_{\sup,K}R^2+\sigma^2}{N}\\ &\le\; (1+\Lambda_k)^{-s}R^2 + \binom{n}{k} C\Lambda_k^{3k/2}\frac{C^2_{\sup,K}R^2+\sigma^2}{N}.\end{aligned}$$

Common cutoff. Balancing bias and variance: $(1+\Lambda_k)^{-s}R^2 \asymp \binom{n}{k}\Lambda_k^{3k/2}/N$. Solving yields the level-k *common cutoff*

$$\Lambda_k \;\asymp\; \max\left\{\Lambda_{\min,k},\; \left(\frac{N}{\binom{n}{k}}\right)^{2/(2s+3k)}\right\}$$

(with $R, \sigma, C_{\sup,K}$ absorbed into the constant), depending only on N, n, k, s and universal constants—*not* on target-specific quantities such as $\|f^*_S\|_{H^s}$. For small N we fix $\Lambda_k = \Lambda_{\min,k}$; this only affects constants in the non-asymptotic regime. Substituting:

$$\sum_{|S|=k}\mathbb{E}\|\widehat{f}_S - f^*_S\|^2_{L^2} \;\le\; C'\left(\frac{\binom{n}{k}}{N}\right)^{2s/(2s+3k)}.$$

Top-order dominance. For $k < K$: since $3k < 3K$, $2s/(2s+3k) > 2s/(2s+3K)$, and with n, K fixed, $\binom{n}{k}$ is a constant, so $(\binom{n}{k}/N)^{2s/(2s+3k)} = o((\binom{n}{K}/N)^{2s/(2s+3K)})$ as $N \to \infty$. The constant term $O(N^{-1})$ is also faster. Hence the $k = K$ level dominates:

$$\mathbb{E}\|\widehat{f}_N - f^*\|^2_{L^2} \;\le\; C\left(\frac{\binom{n}{K}}{N}\right)^{2s/(2s+3K)},$$

where C may depend on n, K, s, R, σ through embedding and mode-count constants, but not on N or f^*.

Lower bound. For each of the $M = \binom{n}{K}$ subsets S_j at order K, let $\{\Phi_{j,l}\}_{l=1}^{m}$ be an orthonormal family of full-support spectrally truncated Peter–Weyl product modes on $\mathrm{SO}(3)^K$ (all K coordinates carrying $\ell_i \ge 1$), embedded into $\mathcal{W}_{S_j} \subset L^2(\mathbf{G})$ (no invariance projection needed). The restriction to full-support modes removes only those with at least one trivial coordinate, whose count is $O(\Lambda^{3(K-1)/2})$; the retained family still has cardinality $m \asymp \Lambda^{3K/2}$ per subset. Modes across different subsets are orthogonal in $L^2(\mathbf{G})$ by Proposition 10 (since $\Phi_{j,l} \in \mathcal{W}_{S_j}$ and $\mathcal{W}_{S_j} \perp \mathcal{W}_{S_{j'}}$ for $j \ne j'$). Set $p = mM$, apply Varshamov–Gilbert (Lemma 28) and Fano (Lemma 27). H^s-membership: $\|\Phi_{j,l}\|^2_{H^s} \le (1+\Lambda)^s$ (direct Sobolev bound on $\mathrm{SO}(3)^K$, no pullback factor). Balance membership and KL constraints: $\Lambda \asymp (NR^2/(M\sigma^2))^{2/(3K+2s)}$. Rate: $\delta^2 \asymp (M/N)^{2s/(2s+3K)} = (\binom{n}{K}/N)^{2s/(2s+3K)}$. □

I Additive Decomposition Proof

The arithmetic in this section is simple, but it should be read together with Theorem 1. The nontrivial content is that all four effective dimensions are minimax-sharp by matching upper and lower bounds. In particular, $D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) = 3(K-1)$ relies on the pure-S pullback (Theorem 18) and the combinatorial lower-bound packing across the $\binom{n}{K}$ top-order subsets (Appendix G). Thus the identity below is a minimax-sharp structural consequence—not merely a definition of effective dimension.

Corollary 36 (Additive decomposition — restated)**.** *Under the assumptions of Theorem 1, and for $2 \le K \le n$, the total minimax effective-dimension reduction from $\mathcal{F}_0$ to $\mathcal{F}_{K,\text{inv}}$ decomposes additively along disjoint axes:*

$$D_{\text{eff}}(\mathcal{F}_0) - D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) \;=\; 3n - 3(K-1) \;=\; \underbrace{3}_{\textit{invariance}} \;+\; \underbrace{3(n-K)}_{\textit{body-order}}.$$

Proof. Theorem 1 gives the four minimax effective dimensions

$$\begin{aligned} D_{\text{eff}}(\mathcal{F}_0) &= 3n, & D_{\text{eff}}(\mathcal{F}_{\text{inv}}) &= 3(n-1),\\ D_{\text{eff}}(\mathcal{F}_K) &= 3K, & D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) &= 3(K-1). \end{aligned}$$

Body-order-then-invariance accounting path ($\mathcal{F}_0 \to \mathcal{F}_K \to \mathcal{F}_{K,\text{inv}}$):

$$\begin{aligned} D_{\text{eff}}(\mathcal{F}_0) - D_{\text{eff}}(\mathcal{F}_K) &= 3n - 3K = 3(n-K),\\ D_{\text{eff}}(\mathcal{F}_K) - D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) &= 3K - 3(K-1) = 3. \end{aligned}$$

Invariance-then-body-order accounting path ($\mathcal{F}_0 \to \mathcal{F}_{\text{inv}} \to \mathcal{F}_{K,\text{inv}}$):

$$\begin{aligned} D_{\text{eff}}(\mathcal{F}_0) - D_{\text{eff}}(\mathcal{F}_{\text{inv}}) &= 3n - 3(n-1) = 3,\\ D_{\text{eff}}(\mathcal{F}_{\text{inv}}) - D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) &= 3(n-1) - 3(K-1) = 3(n-K). \end{aligned}$$

The two paths yield the same total reduction:

$$D_{\text{eff}}(\mathcal{F}_0) - D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) = 3 + 3(n-K) = 3n - 3(K-1).$$

Moreover, each contribution is independent of the other:

$$\begin{aligned} \text{Invariance contribution:}\quad & D_{\text{eff}}(\mathcal{F}_K) - D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) = 3 = D_{\text{eff}}(\mathcal{F}_0) - D_{\text{eff}}(\mathcal{F}_{\text{inv}}),\\ \text{Body-order contribution:}\quad & D_{\text{eff}}(\mathcal{F}_{\text{inv}}) - D_{\text{eff}}(\mathcal{F}_{K,\text{inv}}) = 3(n-K) = D_{\text{eff}}(\mathcal{F}_0) - D_{\text{eff}}(\mathcal{F}_K). \end{aligned}$$

That is, invariance removes a fixed 3 regardless of K, and body-order truncation removes $3(n-K)$ regardless of invariance.

Hence invariance and body-order truncation compose additively at the level of minimax effective dimension: body-order truncation acts *across* ANOVA levels by removing all levels above K, while invariance acts *within* each surviving nonconstant level by quotienting out one global $\mathrm{SO}(3)$ reference frame (a 3-dimensional group). □

Remark 37 (Scope of the additivity)**.** *The statement is an additivity of* effective dimensions*, not of the risk exponents themselves. The minimax exponents are obtained by substituting these effective dimensions into $2s/(2s + D_{\text{eff}})$, which is nonlinear in D_{eff}. The matching lower bound (Appendix G) is essential: without it, the displayed identity would be only an upper-bound dimension accounting, not a minimax-sharp structural decomposition. Upper bounds alone would show only that $\mathcal{F}_{K,\text{inv}}$ can be learned at effective dimension at most $3(K-1)$; the matching lower bound rules out any smaller effective dimension, making the decomposition $3 + 3(n-K)$ exact.*

J Adaptive Estimator

This appendix provides the full proof of Proposition 3 (the fixed-cutoff adaptive oracle inequality stated in the main text). Appendix J.5 records the alternative N-dependent version for completeness and comparison.

J.1 Definitions

Definition 38 (Adaptive estimator via sample splitting, fixed-cutoff setting)**.** *The practitioner fixes,* a priori and independently of N*:*

- *a maximum body-order $K_{\max} \geq 2$;*
- *for each body order $k \in \{2, \dots, K_{\max}\}$, a degree cutoff $\ell_{\max,k} \in \mathbb{Z}_{\geq 0}$. For $|S| = k$ let $\{\phi_{S,j}\}_{j \leq M_{S,k}}$ be the orthonormal basis of Definition 23 for the full-support $P_\Delta^\perp$-range on $\mathrm{SO}(3)^{k-1}$, restricted to per-factor degrees $\sigma_j \leq \ell_{\max,k}$; $M_{S,k}$ is its dimension. Candidate $K' \in \{1, \dots, K_{\max}\}$ uses the dictionary $\mathcal{B}_{K'} := \{T_S \phi_{S,j} : 2 \leq |S| \leq K',\ j \leq M_{S,|S|}\}$, so that $\mathcal{B}_1 = \emptyset$ and $\mathcal{B}_{K'} \subseteq \mathcal{B}_{K'+1}$. For the theoretical counting we associate to $\mathcal{B}_{K'}$ the Casimir envelope*

$$\Lambda_1 := 0, \qquad \Lambda_{K'} := \max_{2 \leq k \leq K'} (k-1)\, \ell_{\max,k}(\ell_{\max,k}+1) \quad (K' \geq 2),$$

which bounds the relative-coordinate Casimir eigenvalues of the basis functions $\phi_{S,j}$ before pullback. The proof uses only that $\mathcal{B}_{K'}$ is finite, fixed independently of N, orthonormal (Lemma 14 within each subset, Proposition 11 across subsets), and that each $\phi_{S,j}$ lies within the Casimir envelope; a joint Casimir retention rule in place of the per-factor rule is an implementation choice (Appendix K).

Given these, define the candidate mode count and the Nikolskii envelope count as

$$M_{K'} := |\mathcal{B}_{K'}| = \sum_{k=2}^{K'} \sum_{|S|=k} M_{S,k}, \qquad \widetilde{M}_{K'} := \sum_{k=2}^{K'} \sum_{|S|=k} \mathsf{N}_{k-1}(\Lambda_{K'}),$$

where $\mathsf{N}_{k-1}(\Lambda)$ is the full spectral mode count on $\mathrm{SO}(3)^{k-1}$ at Casimir $\leq \Lambda$ (Lemma 5). Since each $\phi_{S,j}$ lies in the full Casimir-truncated spectral subspace of $\mathrm{SO}(3)^{k-1}$ at envelope $\Lambda_{K'}$, we have $M_{K'} \leq \widetilde{M}_{K'}$; both are finite constants depending only on n, $K_{\max}$, and the pre-specified cutoffs $\{\ell_{\max,k}\}_{k \leq K'}$, and not on N. For $K' = 1$, both sums are empty, so $M_1 = \widetilde{M}_1 = 0$. The adaptive oracle inequality holds for any such dictionary; in Experiment 4.4, $\ell_{\max,2}$=5 and $\ell_{\max,3}$=1, so the triplet part of $\mathcal{B}_3$ is the $(1,1)$ block on $\mathrm{SO}(3)^2$ restricted to $\mathrm{range}(P_\Delta^\perp)$.

Define, for any $K' \in \{2, \dots, K_{\max}\}$,

$$C_{\sup}(K') := \sqrt{\textstyle\sum_{k=0}^{K'} \binom{n}{k}} \cdot C_{\mathrm{emb}}(s, K'-1),$$

which is finite whenever $s > 3(K'-1)/2$. For the target $f^* \in \mathcal{F}_{K^*,\mathrm{inv}}$, the assumption $s > 3(K^*-1)/2$ of Theorem 1 ensures $\|f^*\|_\infty \leq C_{\sup}(K^*) R$ via Lemma 22. This K^*-specific constant—rather than the uniform $K_{\max}$-version—is all that is required, because the adaptive procedure's analysis depends on $\|f^*\|_\infty$ only through the true K^*.

Given data $(X_\alpha, Y_\alpha)_{\alpha=1}^N$ (for simplicity assume N is even; for odd N replace $N/2$ by $\lfloor N/2 \rfloor, \lceil N/2 \rceil$, which changes only constants):

1. *Split $\{1, \dots, N\}$ into $\mathcal{D}_1, \mathcal{D}_2$ with $|\mathcal{D}_1| = |\mathcal{D}_2| = N/2$.*
2. *For each $K' \in \{1, \dots, K_{\max}\}$, build the spectral projection estimator $\widehat{f}^{(K')}$ from $\mathcal{D}_1$ on the dictionary $\mathcal{B}_{K'}$. For $K' = 1$ the body-order sum below is empty and the candidate reduces to the constant estimator $\widehat{f}^{(1)} = \bar{Y}_{\mathcal{D}_1}$. For $K' \geq 2$:*

$$\widehat{f}^{(K')} = \bar{Y}_{\mathcal{D}_1} + \sum_{k=2}^{K'} \sum_{|S|=k} \sum_{j=1}^{M_{S,k}} \widehat{a}_{S,j}^{(K')}\, T_S \phi_{S,j},$$

   where $\widehat{a}_{S,j}^{(K')} = \frac{2}{N} \sum_{\alpha \in \mathcal{D}_1} Y_\alpha (T_S \phi_{S,j})(X_\alpha) = \frac{2}{N} \sum_{\alpha \in \mathcal{D}_1} Y_\alpha \phi_{S,j}(z_S^{(\alpha)})$ (the two forms coincide by $(T_S \phi_{S,j})(\mathbf{g}) = \phi_{S,j}(z_S(\mathbf{g}))$), the basis is that of Definition 23, truncated per factor at $\ell_{\max,k}$ rather than at an N-dependent level.
3. *Compute holdout risk: $\widehat{R}(K') = \frac{2}{N} \sum_{\alpha \in \mathcal{D}_2} (Y_\alpha - \widehat{f}^{(K')}(X_\alpha))^2$.*
4. *Select $\widehat{K} = \arg\min_{K' \leq K_{\max}} \widehat{R}(K')$.*
5. *Output $\widehat{f}_N = \widehat{f}^{(\widehat{K})}$.*

J.2 Tools

Lemma 39 (Bandlimited Nikolskii inequality on compact groups)**.** *For any function g on $\mathrm{SO}(3)^m$ with spectral support $\lambda_{\boldsymbol{\ell}} \le \Lambda$:*

$$\|g\|_\infty \le \sqrt{\mathsf{N}_m(\Lambda)} \cdot \|g\|_{L^2},$$

where $\mathsf{N}_m(\Lambda) := \sum_{\boldsymbol{\ell}:\lambda_{\boldsymbol{\ell}} \le \Lambda} d_{\boldsymbol{\ell}}^2 \asymp \Lambda^{3m/2}$ is the spectral mode count of Lemma 5.

Proof. By Peter–Weyl expansion and Cauchy–Schwarz: $|g(\mathbf{g})| = |\sum_{\lambda_{\boldsymbol{\ell}} \le \Lambda} d_{\boldsymbol{\ell}} \mathrm{tr}(\widehat{g}(\boldsymbol{\ell}) D^{\boldsymbol{\ell}}(\mathbf{g}))| \le \sum d_{\boldsymbol{\ell}} \|\widehat{g}(\boldsymbol{\ell})\|_{\mathrm{HS}} \|D^{\boldsymbol{\ell}}(\mathbf{g})\|_{\mathrm{HS}} = \sum d_{\boldsymbol{\ell}}^{3/2} \|\widehat{g}(\boldsymbol{\ell})\|_{\mathrm{HS}}$. By Cauchy–Schwarz: $\le (\sum d_{\boldsymbol{\ell}}^2)^{1/2} (\sum d_{\boldsymbol{\ell}} \|\widehat{g}(\boldsymbol{\ell})\|_{\mathrm{HS}}^2)^{1/2} = \sqrt{\mathsf{N}_m(\Lambda)} \cdot \|g\|_{L^2}$. □

Lemma 40 (Candidate estimator sup-norm control, fixed cutoff)**.** *Under the fixed-cutoff assumption of Definition 38, for a target $f^* \in \mathcal{F}_{K^*,\mathrm{inv}}$ satisfying the smoothness assumption of Theorem 1 (namely $s > 3(K^*-1)/2$), and for each $K' \le K_{\max}$ and all $N \ge 1$:*

$$\mathbb{E}\big[\|\widehat{f}^{(K')}\|_\infty^2\big] \;\le\; B_{K'}^2,$$

where

$$B_{K'}^2 := 2\big((C_{\sup}(K^*))^2 R^2 + \sigma^2\big) + 2\widetilde{M}_{K'}\big(R^2 + 2M_{K'}((C_{\sup}(K^*))^2 R^2 + \sigma^2)\big)$$

depends only on $n, K_{\max}, K^*, s, R, \sigma$ and the pre-specified cutoffs $\{\ell_{\max,k}\}$ (via $M_{K'}$ and $\widetilde{M}_{K'}$), and in particular does not grow with N.

Proof. Sup-norm bound via full spectral Nikolskii. Each component $\widehat{f}_S = \sum_j \widehat{a}_{S,j}^{(K')} T_S \phi_{S,j}$ lies in $\mathcal{V}_S$ via the pullback, and $\|\widehat{f}_S\|_{\infty,\mathbf{G}} = \|\phi\|_{\infty,\mathrm{SO}(3)^{k-1}}$ where $\phi = \sum_j \widehat{a}_{S,j}^{(K')} \phi_{S,j}$ is its pre-image (since $z_S : \mathbf{G} \to \mathrm{SO}(3)^{k-1}$ is surjective). The pre-image ϕ has Peter–Weyl spectral support contained in $\{\lambda_{\boldsymbol{\sigma}} \le \Lambda_{K'}\}$ on $\mathrm{SO}(3)^{k-1}$ (as a subspace of the full Casimir-truncated spectral subspace, which *is* product-left-$\mathrm{SO}(3)^{k-1}$-invariant; note the pure-S subspace itself is not product-invariant because $P_\Delta^\perp$ is defined via the diagonal action only). Applying Lemma 39:

$$\|\widehat{f}_S\|_\infty \le \sqrt{\mathsf{N}_{k-1}(\Lambda_{K'})}\, \|\widehat{f}_S\|_{L^2}.$$

By triangle inequality and Cauchy–Schwarz over subsets and body-order levels:

$$\|\widehat{f}^{(K')} - \bar{Y}\|_\infty \le \sum_{k,S} \sqrt{\mathsf{N}_{k-1}(\Lambda_{K'})}\, \|\widehat{f}_S\|_{L^2} \le \sqrt{\widetilde{M}_{K'}} \cdot \sqrt{\sum_{S,j} \widehat{a}_{S,j}^2},$$

where the last step uses Cauchy–Schwarz across the (S,k) sum with weights $\mathsf{N}_{k-1}(\Lambda_{K'})$ summing to $\widetilde{M}_{K'}$, and $\sum_S \|\widehat{f}_S\|_{L^2}^2 = \sum_{S,j} \widehat{a}_{S,j}^2$ by orthonormality of $\{T_S \phi_{S,j}\}$ (Lemma 14 and Proposition 11).

Coefficient expectation via Bessel. Since $\{T_S \phi_{S,j}\}_{S,j}$ forms an orthonormal system in $L^2(\mathbf{G})$ (Lemma 14 and Proposition 11), Bessel's inequality gives

$$\sum_{S,j} \langle f^*,\, T_S \phi_{S,j} \rangle_{L^2(\mathbf{G})}^2 \;\le\; \|f^*\|_{L^2(\mathbf{G})}^2 \;\le\; R^2.$$

By the unbiasedness calculation (Proposition 24), $\mathbb{E}[\widehat{a}_{S,j}^{(K')}] = \langle f^*, T_S \phi_{S,j} \rangle_{L^2(\mathbf{G})}$. Combining with Proposition 25 (adapted to $|\mathcal{D}_1| = N/2$), the per-coefficient variance satisfies $\mathrm{Var}(\widehat{a}_{S,j}^{(K')}) \le 2((C_{\sup}(K^*))^2 R^2 + \sigma^2)/N$, so

$$\mathbb{E}\big[\sum_{S,j} (\widehat{a}_{S,j}^{(K')})^2\big] = \sum_{S,j} \big(\mathbb{E}[\widehat{a}_{S,j}^{(K')}]^2 + \mathrm{Var}(\widehat{a}_{S,j}^{(K')})\big) \le R^2 + \frac{2M_{K'}((C_{\sup}(K^*))^2 R^2 + \sigma^2)}{N}.$$

For $N \ge 1$, the RHS is bounded by $R^2 + 2M_{K'}((C_{\sup}(K^*))^2 R^2 + \sigma^2)$.

Combining. Therefore $\mathbb{E}[\|\widehat{f}^{(K')} - \bar{Y}\|_\infty^2] \le \widetilde{M}_{K'}\big(R^2 + 2M_{K'}((C_{\sup}(K^*))^2 R^2 + \sigma^2)\big)$. Combining with $\mathbb{E}[\bar{Y}^2] \le (C_{\sup}(K^*))^2 R^2 + \sigma^2$ via $(a+b)^2 \le 2a^2 + 2b^2$ yields the stated bound. □

J.3 Main proof

Proposition 41 (Adaptive oracle inequality, fixed-cutoff family — restated)**.** *Under the Gaussian regression model of Section 3.1 ($Y_\alpha = f^*(X_\alpha) + \xi_\alpha$, $\xi_\alpha \sim \mathcal{N}(0,\sigma^2)$ i.i.d., $X_\alpha \sim \mu^{\otimes n}$), the fixed-cutoff assumption of Definition 38, and the smoothness assumption of Theorem 1 for $\mathcal{F}_{K^*,\mathrm{inv}}$ (namely $s > 3(K^*-1)/2$, which ensures $\|f^*\|_\infty < \infty$ via Lemma 22), for all $f^* \in \mathcal{F}_{K^*,\mathrm{inv}}$ with $2 \le K^* \le K_{\max}$:*

$$\mathbb{E}\big[\|\widehat{f}_N - f^*\|_{L^2}^2\big] \;\le\; \min_{K' \le K_{\max}} R_N(K') \;+\; \tau_N,$$

where $R_N(K') := \mathbb{E}\|\widehat{f}^{(K')} - f^*\|_{L^2}^2$ is the risk of the K'-candidate and

$$\tau_N \;\le\; C\sqrt{\frac{\log(2K_{\max}N)}{N}},$$

with C possibly depending on $n, K_{\max}, s, R, \sigma$, the fixed cutoffs $\{\ell_{\max,k}\}_{k=2}^{K_{\max}}$ (through $\max_{K'} B_{K'}^2$ from Lemma 40), and the true-class sup-norm constant $C_{\sup}(K^*)$, but not on N or on the particular target $f^* \in \mathcal{F}_{K^*,\mathrm{inv}}$. In particular, $\tau_N = O(\sqrt{\log N/N}) \to 0$, and no smoothness condition beyond that of Theorem 1 for $\mathcal{F}_{K^*,\mathrm{inv}}$ is required, in contrast to Proposition 43 (Appendix J.5), which requires $s > 3(K_{\max}-1)/2$ uniformly over candidates. The benchmark $\min_{K'} R_N(K')$ is the in-family oracle over the pre-specified candidates; the inequality does not assert attainment of the minimax rate (Remark 42).

Proof. Throughout, $\mathcal{D}_1$ is the fitting split and $\mathcal{D}_2$ the holdout split; candidate estimators $\widehat{f}^{(K')}$ are $\mathcal{D}_1$-measurable.

Step 1: Per-candidate risk as benchmark. $R_N(K') := \mathbb{E}_{\mathcal{D}_1}[\|\widehat{f}^{(K')} - f^*\|_{L^2}^2]$ is the integrated in-family risk. Unlike the N-dependent analysis of Appendix J.5, we do not decompose $R_N(K')$ into approximation and estimation terms against a projected target: the benchmark incorporates any approximation error of the fixed dictionaries, and the proof bounds only the excess risk from selection. Because the dictionaries are finite and fixed, the expected squared sup-norm of each candidate is bounded uniformly in N (Lemma 40), which is what removes the condition $s > 3(K_{\max}-1)/2$ of the N-dependent version.

Step 2: Holdout risk decomposition. Let $R_{\mathrm{true}}(K') := \|\widehat{f}^{(K')} - f^*\|_{L^2}^2$, which is $\mathcal{D}_1$-measurable. Given $\mathcal{D}_1$: $\mathbb{E}_{\mathcal{D}_2}[\widehat{R}(K') \mid \mathcal{D}_1] = R_{\mathrm{true}}(K') + \sigma^2$ by standard bias-variance decomposition of the squared prediction loss.

Step 3: Concentration via Bernstein. Condition on $\mathcal{D}_1$. For each K', the summands $Z_\alpha^{(K')} := (Y_\alpha - \widehat{f}^{(K')}(X_\alpha))^2$, $\alpha \in \mathcal{D}_2$, are i.i.d. Since $Y_\alpha - \widehat{f}^{(K')}(X_\alpha) = (f^*(X_\alpha) - \widehat{f}^{(K')}(X_\alpha)) + \xi_\alpha$ is bounded in magnitude by $\|f^*\|_\infty + \|\widehat{f}^{(K')}\|_\infty$ plus a $\mathcal{N}(0,\sigma^2)$ term, it is sub-Gaussian with parameter $\le C_1(\|f^*\|_\infty + \|\widehat{f}^{(K')}\|_\infty + \sigma)$; its square $Z_\alpha^{(K')}$ is therefore sub-exponential [34], with

$$\|Z_\alpha^{(K')} - \mathbb{E}[Z_\alpha^{(K')}]\|_{\psi_1} \le C_2(\|f^*\|_\infty + \|\widehat{f}^{(K')}\|_\infty + \sigma)^2.$$

Define the uniform envelope

$$b_{\max}(\mathcal{D}_1) := C_{\sup}(K^*)R + \max_{K' \le K_{\max}} \|\widehat{f}^{(K')}\|_\infty + \sigma,$$

so that $\|Z_\alpha^{(K')} - \mathbb{E}[Z_\alpha^{(K')}]\|_{\psi_1} \le C_2\, b_{\max}^2$ uniformly in K'. By Bernstein's inequality, for each K' and $t > 0$:

$$\Pr\big(|\widehat{R}(K') - (R_{\mathrm{true}}(K') + \sigma^2)| > t \;\big|\; \mathcal{D}_1\big) \;\le\; 2\exp\big(-cN\min(t^2/b_{\max}^4, t/b_{\max}^2)\big).$$

Step 4: Union bound. Set $t_N(\mathcal{D}_1) := C_3\, b_{\max}(\mathcal{D}_1)^2\sqrt{\log(2K_{\max}N)/N}$. A union bound over the $K_{\max}$ candidates gives

$$\Pr\big(\exists\, K' \le K_{\max} : |\widehat{R}(K') - (R_{\mathrm{true}}(K') + \sigma^2)| > t_N \;\big|\; \mathcal{D}_1\big) \le 1/N.$$

Call the good event $\mathcal{E}(\mathcal{D}_1)$; note $\mathcal{E}$ is $\mathcal{D}_2$-measurable given $\mathcal{D}_1$.

Step 5: Oracle comparison on $\mathcal{E}$. On $\mathcal{E}$, all candidate holdout risks are within t_N of their conditional means. In particular, for any K':

$$R_{\mathrm{true}}(\widehat{K}) = (\widehat{R}(\widehat{K}) - \sigma^2) + \underbrace{(R_{\mathrm{true}}(\widehat{K}) + \sigma^2 - \widehat{R}(\widehat{K}))}_{\le t_N} \le \widehat{R}(K') - \sigma^2 + t_N$$

(by the selection rule $\widehat{R}(\widehat{K}) \le \widehat{R}(K')$), and similarly $\widehat{R}(K') \le R_{\mathrm{true}}(K') + \sigma^2 + t_N$, so

$$R_{\mathrm{true}}(\widehat{K}) \le R_{\mathrm{true}}(K') + 2t_N \quad \text{on } \mathcal{E}, \text{ for every } K'.$$

In particular, $R_{\mathrm{true}}(\widehat{K}) \le \min_{K'} R_{\mathrm{true}}(K') + 2t_N$ on $\mathcal{E}$.

On $\mathcal{E}^c$, we use the crude bound $R_{\mathrm{true}}(\widehat{K}) = \|\widehat{f}^{(\widehat{K})} - f^*\|_{L^2}^2 \le \|\widehat{f}^{(\widehat{K})} - f^*\|_\infty^2 \le b_{\max}^2$, since $b_{\max}$ dominates $\|\widehat{f}^{(\widehat{K})}\|_\infty + \|f^*\|_\infty$.

Step 6: Integration. Taking expectation over $(\mathcal{D}_1, \mathcal{D}_2)$:

$$\begin{aligned}\mathbb{E}[R_{\mathrm{true}}(\widehat{K})] &= \mathbb{E}[R_{\mathrm{true}}(\widehat{K})\,\mathbb{1}\{\mathcal{E}\}] + \mathbb{E}[R_{\mathrm{true}}(\widehat{K})\,\mathbb{1}\{\mathcal{E}^c\}] \\ &\le \mathbb{E}[\min_{K'} R_{\mathrm{true}}(K') + 2t_N] + \mathbb{E}[b_{\max}^2\,\mathbb{1}\{\mathcal{E}^c\}].\end{aligned}$$

For the first expectation, $\mathbb{E}[\min_{K'} R_{\mathrm{true}}(K')] \le \min_{K'} \mathbb{E}[R_{\mathrm{true}}(K')] = \min_{K'} R_N(K')$. The second expectation is handled by the tower rule: since $b_{\max}$ is $\mathcal{D}_1$-measurable and $\Pr(\mathcal{E}^c \mid \mathcal{D}_1) \le 1/N$,

$$\mathbb{E}[b_{\max}^2\,\mathbb{1}\{\mathcal{E}^c\}] = \mathbb{E}_{\mathcal{D}_1}\big[b_{\max}^2 \cdot \Pr(\mathcal{E}^c \mid \mathcal{D}_1)\big] \le \mathbb{E}[b_{\max}^2]/N.$$

Step 7: N-uniform envelope bound. Here is where the fixed-cutoff structure enters decisively: unlike the N-dependent regime (Appendix J.5), $M_{K'}$ is a fixed constant, so by Lemma 40

$$\mathbb{E}[\max_{K' \le K_{\max}} \|\widehat{f}^{(K')}\|_\infty^2] \le \sum_{K'=1}^{K_{\max}} \mathbb{E}[\|\widehat{f}^{(K')}\|_\infty^2] \le K_{\max} \max_{K'} B_{K'}^2 =: B^2,$$

with B independent of N. By $(a+b+c)^2 \le 3(a^2+b^2+c^2)$:

$$\mathbb{E}[b_{\max}^2] \le 3\big((C_{\sup}(K^*))^2 R^2 + B^2 + \sigma^2\big) =: \bar{B}^2,$$

a constant independent of N.

Step 8: Conclude. Therefore:

$$\begin{aligned}\mathbb{E}[t_N] &\le C_3\,\bar{B}^2\,\sqrt{\log(2K_{\max}N)/N}, \\ \mathbb{E}[b_{\max}^2\,\mathbb{1}\{\mathcal{E}^c\}] &\le \bar{B}^2/N.\end{aligned}$$

Combining, and noting $\bar{B}^2/N \le \bar{B}^2\sqrt{\log(2K_{\max}N)/N}$ for $N \ge 1$:

$$\tau_N := 2\mathbb{E}[t_N] + \bar{B}^2/N \;\le\; C\sqrt{\log(2K_{\max}N)/N},$$

with $C := (2C_3 + 1)\bar{B}^2$. The bound is $O(\sqrt{\log N/N}) \to 0$ as $N \to \infty$, completing the proof. □

J.4 Remark and discussion

Remark 42 (Scope of the fixed-family oracle)**.** *Proposition 3 benchmarks the adaptive estimator against $\min_{K'} R_N(K')$, the best risk within the pre-specified fixed-cutoff family. This benchmark and the minimax rate $r_N(K^*) = (\binom{n}{K^*}/N)^{2s/(2s+3(K^*-1))}$ of Theorem 1 are different objects: a fixed dictionary can incur approximation error unless the target lies in the candidate's finite-dimensional approximation space, so the in-family oracle is not in general of order $r_N(K^*)$. The candidate at K^* attains $r_N(K^*)$ under the level-wise N-dependent cutoffs of Theorem 26, which leaves the fixed-cutoff setting; Appendix J.5 records the corresponding adaptive statement. Selection over a grid of cutoffs, by Lepski's method or penalization, is left open.*

J.5 Comparison with N-dependent cutoff adaptivity

For completeness, we record an alternative scheme in which each candidate's Casimir threshold is tuned jointly with N. This version achieves sharper asymptotic benchmarks, at the cost of a stronger smoothness assumption.

Proposition 43 (N-dependent cutoff version)**.** *Suppose $s > 3(K_{\max}-1)/2$. For $K'=1$ the candidate is the constant estimator, so $M_1^{(N)} = \widetilde{M}_1^{(N)} = 0$. For $2 \le K' \le K_{\max}$, let candidate K' retain all pure modes of body orders $k \le K'$ with relative-coordinate Casimir eigenvalue at most $\Lambda_{K'}^{(N)} \asymp N^{2/(2s+3(K'-1))}$, so that its mode count is $M_{K'}^{(N)} \asymp N^{3(K'-1)/(2s+3(K'-1))}$. Then the analogous sample-splitting procedure, run as in Definition 38 on these N-dependent dictionaries, satisfies, for all $f^* \in \mathcal{F}_{K^*,\mathrm{inv}}$ with $2 \le K^* \le K_{\max}$:*

$$\mathbb{E}\|\widehat{f}_N - f^*\|^2 \;\le\; C_1\, r_N(K^*) \;+\; O\big(N^{-\gamma}\sqrt{\log N}\big), \qquad \gamma = \frac{2s-3(K_{\max}-1)}{2(2s+3(K_{\max}-1))} > 0,$$

where $r_N(K^*) = \big(\binom{n}{K^*}/N\big)^{2s/(2s+3(K^*-1))}$ is the minimax rate on $\mathcal{F}_{K^*,\mathrm{inv}}$.

Proof sketch. The argument follows the template of Proposition 41 through Step 6. Two modifications enter.

(a) Misspecification handling in Step 1. Define $\mathcal{L}_{K'} := \bigoplus_{k\le K'} \bigoplus_{|S|=k} \mathcal{V}_S$, the invariant-ANOVA subspace of body-order $\le K'$; $\Pi_{K'}$ denotes L^2-orthogonal projection onto $\mathcal{L}_{K'}$ (well-defined by Proposition 11). For $K' < K^*$, the candidate $\widehat{f}^{(K')}$ is misspecified ($f^* \notin \mathcal{L}_{K'}$ in general), and the approximation bias $\|\Pi_{K'}f^* - f^*\|^2$ may be a non-vanishing constant that is *not* bounded by the minimax rate. This is harmless for the oracle comparison, because

$$\min_{K'\le K_{\max}} R_N(K') \;\le\; R_N(K^*),$$

and the $K' = K^*$ candidate is well-specified ($\Pi_{K^*}f^* = f^*$, approximation bias vanishes). Let $D_* := 3(K^*-1)$. At the common threshold $\Lambda_{K^*}^{(N)}$, the squared bias of $\widehat{f}^{(K^*)}$ is $O((\Lambda_{K^*}^{(N)})^{-s})$ and its variance is $O((\Lambda_{K^*}^{(N)})^{D_*/2}/N)$ by the bias–variance bounds of Appendix F with sample size $N/2$; both are $O(r_N(K^*))$, so $R_N(K^*) \le C_0 r_N(K^*)$, whence

$$\min_{K'} R_N(K') \;\le\; C_0 r_N(K^*).$$

(b) N-dependent envelope bound in Step 7. Repeating the argument of Lemma 40 with the N-dependent dictionaries gives

$$\mathbb{E}[\|\widehat{f}^{(K')}\|_\infty^2] = O\Big(1 + \widetilde{M}_{K'}^{(N)}\big(1 + M_{K'}^{(N)}/N\big)\Big) = O(1 + \widetilde{M}_{K'}^{(N)})$$

(since $M_{K'}^{(N)}/N \le 1$ at the optimal cutoff under $s > 0$), where $\widetilde{M}_{K'}^{(N)}$ is the Nikolskii envelope count at threshold $\Lambda_{K'}^{(N)}$. For $K' \ge 2$, $\widetilde{M}_{K'}^{(N)} \asymp M_{K'}^{(N)} \asymp (\Lambda_{K'}^{(N)})^{3(K'-1)/2} \asymp N^{3(K'-1)/(2s+3(K'-1))}$ (both pure-S count and full spectral count scale with the same order in $\Lambda_{K'}^{(N)}$, differing only by a constant factor). Therefore $\mathbb{E}[b_{\max}^2] = O(1 + \max_{K'} \widetilde{M}_{K'}^{(N)}) = O(N^{D/(2s+D)})$ where $D := 3(K_{\max}-1)$.

Bernstein bound in Step 8. $\mathbb{E}[t_N] = O\big(N^{D/(2s+D)} \cdot \sqrt{\log N/N}\big) = O\big(N^{D/(2s+D)-1/2}\sqrt{\log N}\big)$. Under $s > D/2$: $D/(2s+D) < D/(D+D) = 1/2$, so the exponent $D/(2s+D) - 1/2 = -\gamma < 0$ with $\gamma = (2s-D)/(2(2s+D)) > 0$, giving $\tau_N = O(N^{-\gamma}\sqrt{\log N}) \to 0$.

Combining (a) and (b) via Step 6 of the main proof: $\mathbb{E}\|\widehat{f}_N - f^*\|^2 \le \min_{K'} R_N(K') + \tau_N \le C_0 r_N(K^*) + O(N^{-\gamma}\sqrt{\log N})$, giving the claim with $C_1 = C_0$. □

Comparison of the two versions. Proposition 3 (fixed-cutoff) and Proposition 43 (N-dependent) differ in three respects:

- *Benchmark*: fixed-cutoff compares the adaptive risk against $\min_{K'} R_N(K')$, the best risk within the pre-specified family; N-dependent compares against $r_N(K^*)$, the minimax rate.

- *Smoothness requirement*: fixed-cutoff requires only $s > 3(K^*-1)/2$, i.e., the same smoothness condition Theorem 1 imposes on $\mathcal{F}_{K^*,\mathrm{inv}}$ for the true f^*; N-dependent additionally requires this condition to hold *uniformly* over all candidate body-orders $K' \le K_{\max}$, which is equivalent to $s > 3(K_{\max}-1)/2$ and becomes increasingly stringent as $K_{\max}$ grows (e.g., $s > 3$ for $K_{\max} = 3$, $s > 4.5$ for $K_{\max} = 4$).
- *Cutoff tuning*: fixed-cutoff uses cutoffs pre-specified by the practitioner; N-dependent requires thresholds tuned to each N, which requires knowing s.

We focus on the fixed-cutoff version (Proposition 3) in the main text because the experimental setup is that regime, with $\{\ell_{\max,k}\}$ chosen by computational feasibility and held fixed across all tested N (Section 4.4), and because its smoothness requirement is the one Theorem 1 already imposes.

K Experimental Details

This section specifies estimator implementations, ground truth constructions, risk evaluation protocol, cross-experiment parameter values, and compute resources.

K.1 Estimator implementations

All estimators operate in the realified Wigner-D basis and share a common structural template: each estimator fixes a finite retained dictionary of Peter–Weyl modes and computes empirical Fourier coefficients on the training data.

Basis convention. In implementation we use the real orthonormal basis obtained from the real and imaginary parts of the Wigner-D matrix coefficients (consistent with the theoretical convention declared in Appendix F). The displayed $D^{\ell}_{mm'}$ notation in the coefficient formulas below is shorthand for this realified basis; all estimator coefficients $\widehat{a}$ are real-valued, matching real-valued observations $Y_\alpha \in \mathbb{R}$.

Stabilized spectral implementation (Sobolev-weighted shrinkage). The minimax proofs in Appendices F–G analyze a hard-cutoff spectral projection for clarity of argument. For numerical stability in the finite-sample regime, the code applies a Sobolev-weighted Tikhonov shrinkage to each retained coefficient:

$$\widehat{a}_q \;\leftarrow\; \frac{1}{1+\mu(1+\lambda_q)^s}\,\widehat{a}_q, \qquad \mu = (1+\Lambda)^{-s},$$

where λ_q is the Casimir eigenvalue of mode q and $\Lambda \asymp N^{2/(2s+D)}$ is the class-specific N-dependent optimal bias–variance cutoff scale. This is a standard stabilized version of spectral projection: on Sobolev balls, hard spectral truncation and Sobolev-weighted Tikhonov regularization have the same bias–variance scaling up to constant factors [35, 32], so the experiments should be interpreted as stabilized implementations of the same structural family rather than literal implementations of the proof estimator.

Experiment-specific shrinkage scales. For the $\mathcal{F}_{K,\mathrm{inv}}$ estimator in Experiment 1 and its s=3 variant (Appendix L), the implemented parameter is $\mu = (1+\Lambda^*)^{-s}/\sqrt{\binom{n}{2}}$, where $\Lambda^* = (N/\binom{n}{2})^{2/(2s+3)}$. In the K=3 study (Appendix N), both estimator families use $\mu_k = (1+\Lambda_k^*)^{-s}/\sqrt{\binom{n}{k}}$ at each retained nonconstant body order, where $\Lambda_k^* = (N/\binom{n}{k})^{2/(2s+D_k)}$, $D_k = 3(k-1)$ for $\mathcal{F}_{K,\mathrm{inv}}$ ($k = 2, 3$), and $D_k = 3k$ for $\mathcal{F}_K$ ($k = 1, 2, 3$). Experiments 2–5, including the Experiment 5 robustness variants, use the base shrinkage convention without this additional factor.

The key implementation difference across classes is the *coordinate system* and the *effective ambient space* in which the Fourier expansion is performed:

- $\mathcal{F}_{K,\mathrm{inv}}$: For each pair (i,j), compute relative orientation $r_{ij} = g_i^{-1}g_j \in \mathrm{SO}(3)$, expand in the Wigner-$D$ basis on $\mathrm{SO}(3)$ up to $\ell_{\max}$, estimate coefficients $\widehat{a}_{\ell,m,m'} = \frac{1}{N}\sum_{\alpha=1}^{N} Y_\alpha \sqrt{d_\ell}\, D^{\ell}_{mm'}(r_{ij}^{(\alpha)})$, and sum across all $\binom{n}{2}$ pairs. Each pair contributes $\sum_{\ell=1}^{\ell_{\max}}(2\ell+1)^2$ modes; the total mode count is $\binom{n}{2}\sum_{\ell=1}^{\ell_{\max}}(2\ell+1)^2$.

- $\mathcal{F}_K$: For each pair (i, j), expand directly in the Wigner-D product basis on $\mathrm{SO}(3)^2$ (absolute coordinates) with per-factor cutoff $\ell_{\max,2b}$ for the two-body terms, plus separate one-body terms for each coordinate at cutoff $\ell_{\max,1b}$ (both specified in Table 3). The 2-body mode count per pair is $(\sum_{\ell=1}^{\ell_{\max,2b}}(2\ell+1)^2)^2$, significantly larger than the invariant case; summing across $\binom{n}{2}$ pairs and adding $n \cdot \sum_{\ell=1}^{\ell_{\max,1b}}(2\ell+1)^2$ one-body modes gives the total.
- $\mathcal{F}_{\mathrm{inv}}$: Expand in the Wigner-$D$ product basis on $\mathrm{SO}(3)^{n-1}$ using all relative coordinates $(g_1^{-1}g_2, \ldots, g_1^{-1}g_n)$, with mode retention under joint Casimir truncation (see *Convention by estimator family* below).
- $\mathcal{F}_0$: Expand in the Wigner-D product basis on $\mathrm{SO}(3)^n$ (absolute coordinates), with mode retention under joint Casimir truncation (see *Convention by estimator family* below).

Cutoff selection and truncation convention. The cutoff $\ell_{\max}$ is chosen so that the total number of retained basis functions remains computationally tractable (ranging from $\sim 3 \times 10^3$ modes for low-dimensional classes at small n up to $\sim 10^5$ modes for $\mathcal{F}_0$ at n=10) and numerically stable. Because a naïve per-factor mode count would grow as $\ell_{\max}^{3m}$ where m is the number of $\mathrm{SO}(3)$ factors, higher-dimensional classes ($\mathcal{F}_{\mathrm{inv}}$, $\mathcal{F}_0$) use smaller $\ell_{\max}$ *and* switch to joint Casimir truncation (see *Convention by estimator family* below) to avoid exceeding the sample size. This reflects the theoretical scaling: the effective dimension governs both the minimax rate and the computational cost of the estimator.

Convention by estimator family. The theoretical analysis is stated in terms of joint Casimir thresholds $\lambda_{\boldsymbol{\ell}} \le \Lambda$. In the experiments, the implementation convention depends on the estimator family:

- For the high-dimensional product-domain estimators $\mathcal{F}_{\mathrm{inv}}$ on $\mathrm{SO}(3)^{n-1}$ and $\mathcal{F}_0$ on $\mathrm{SO}(3)^n$, we use a joint Casimir truncation $\lambda_{\boldsymbol{\ell}} = \sum_i \ell_i(\ell_i+1) \le \Lambda_{\mathrm{cut}}$, because a naïve per-factor tensor cutoff "each $\ell_i \le \ell_{\max}$" would produce a mode count $\left(\sum_{\ell=0}^{\ell_{\max}}(2\ell+1)^2\right)^m$ that is computationally infeasible at moderate n (e.g., $35^9 \approx 8 \times 10^{13}$ for $\mathcal{F}_{\mathrm{inv}}$ at n=10, $\ell_{\max}$=2). Let $\Lambda^*(N) = (N/M)^{2/(2s+D)}$, with $M = 1$ for these two classes. In Experiment 1, we use $\Lambda_{\mathrm{cut}}(N) = \max\{3\Lambda^*(N), 6\}$. In Experiment 2, the joint cutoff is computed at $N_{\max}$ and held fixed across sample sizes. In Experiments 3 and 5, we fix $\Lambda_{\mathrm{cut}} = 6$. In all four experiments, the shrinkage parameter $\mu(N) = (1 + \Lambda^*(N))^{-s}$ is updated with N. The listed $\ell_{\max}$ values for $\mathcal{F}_{\mathrm{inv}}$ and $\mathcal{F}_0$ in Table 3 cap the representation degree of each factor; the joint Casimir threshold Λ_{cut} is set separately as described above.
- For the pair- and triplet-separable estimators $\mathcal{F}_{K,\mathrm{inv}}$ and $\mathcal{F}_K$ (and the body-order-K' adaptive candidates in Experiment 4.4), we use fixed per-factor cutoffs within each pair/triplet expansion, independently of N. Thus $\mathcal{F}_{K,\mathrm{inv}}$ retains relative modes $1 \le \ell \le \ell_{\max}$ on $\mathrm{SO}(3)$ per pair; $\mathcal{F}_K$ retains absolute pair modes $1 \le \ell_1, \ell_2 \le \ell_{\max}$ on $\mathrm{SO}(3)^2$ per pair (plus the specified 1-body cutoff); and the adaptive K'=3 candidate retains, per triplet, the full-support relative triplet modes with all $\sigma_j \le \ell_{\max,3}$ on $\mathrm{SO}(3)^2$, restricted to $\mathrm{range}(P_\Delta^\perp)$ following Definition 38.

For $\mathcal{F}_{K,\mathrm{inv}}$, the pair relative domain consists of a single $\mathrm{SO}(3)$ factor, so the degree cutoff $\ell \le \ell_{\max}$ is equivalent to the Casimir cutoff $\ell(\ell+1) \le \ell_{\max}(\ell_{\max}+1)$. For $\mathcal{F}_K$ (pair absolute on $\mathrm{SO}(3)^2$) and for the $K' = 3$ triplet candidate (relative on $\mathrm{SO}(3)^2$), the two conventions differ by constant-level finite-dimensional choices at fixed $\ell_{\max}$ but preserve the same polynomial mode-count scaling; this leaves the qualitative structural hierarchy and slope comparisons unchanged across cutoff conventions, although finite-dimensional constants may differ. The adaptive oracle inequality (Proposition 3) requires only that each candidate dictionary is finite and fixed independently of N, which holds under either retention rule.

Risk evaluation. Training labels are generated as $Y_\alpha = f^*(X_\alpha) + \xi_\alpha$ with $\xi_\alpha \sim \mathcal{N}(0, \sigma^2)$ i.i.d., while reported test risks are computed against the clean target f^*. All risks are estimated as $\widehat{R} = \frac{1}{N_{\mathrm{test}}}\sum_{\alpha=1}^{N_{\mathrm{test}}}(\widehat{f}(X_\alpha^{\mathrm{test}}) - f^*(X_\alpha^{\mathrm{test}}))^2$ on a held-out test set of $N_{\mathrm{test}} = 5{,}000$ fresh samples (drawn independently from the training data). For log-log slope extraction (Experiments 1–2), we use ordinary least squares on $\log_{10}(\widehat{R})$ vs. $\log_{10}(N)$. Each configuration is run for $M_{\mathrm{rep}} \in \{20, 30\}$ independent repetitions (Table 3); in all log-log risk-vs-N learning-curve panels, curves show the mean across repetitions with error bars giving one standard deviation. These error bars are visually most prominent in the small-N and transition regimes (e.g., Figures 4 and 5); at larger N, the run-to-run standard deviation becomes small relative to the mean on the log-log scale, so error bars

may appear close to the marker size in some panels (e.g., Figures 2 and 3). Secondary panels that show derived aggregates (rate-exponent bar charts, regression slopes as a function of n, risk ratios, and selection-frequency distributions) report point estimates without additional uncertainty bars.

K.2 Ground truth construction

- **Experiment 1** (per-class rate validation): Each class uses its own ground truth, constructed by sampling random Peter–Weyl coefficients with Sobolev-type spectral decay. Concretely, for each retained Peter–Weyl matrix coefficient

$$\psi_q = (2\ell+1)\, D^{\ell}_{mm'}$$

(or the analogous tensor-product coefficient on a pair, triplet, or product domain) with Casimir eigenvalue λ_q, we draw

$$c_q = (1+\lambda_q)^{-(s+1)/2}\, \xi_q, \qquad \xi_q \sim \mathcal{N}_{\mathbb{C}}(0,1),$$

so that $\mathbb{E}|c_q|^2 = (1+\lambda_q)^{-(s+1)}$, and set

$$f^* = \mathrm{Re}\left(\sum_q c_q\, \psi_q \right),$$

followed by rescaling so that $\mathrm{Var}(f^*(X)) \approx (0.1)^2$ on a held-out sample. Coefficients are drawn up to a ground-truth truncation level $\ell_{\max,\mathrm{GT}} > \ell_{\max}$ to ensure the target carries energy beyond the estimator's retained dictionary. We use "Sobolev-type decay" to refer to this finite-truncated spectral decay controlled by the parameter s; the generator is used as a controlled synthetic target for testing the structural hierarchy and is not a formal infinite-dimensional Sobolev prior. The basis $\{\psi_q\}$ per class is: pairwise relative modes on $\mathrm{SO}(3)$ for $\mathcal{F}_{K,\mathrm{inv}}$; absolute one-body modes on $\mathrm{SO}(3)$ plus two-body modes on $\mathrm{SO}(3)^2$ for $\mathcal{F}_K$; product modes on $\mathrm{SO}(3)^{n-1}$ for $\mathcal{F}_{\mathrm{inv}}$; and product modes on $\mathrm{SO}(3)^n$ for $\mathcal{F}_0$.

- **Experiment 2** (n-dependence): Same per-class construction as Experiment 1, repeated for each $n \in \{3,4,5,6,7,8\}$. A fresh ground truth is drawn for each (n, class) pair.

- **Experiment 3** (four-way hierarchy): Three ground truths $f^* \in \mathcal{F}_{K,\mathrm{inv}}$ (seeds 42, 43, 44) are generated as in the $\mathcal{F}_{K,\mathrm{inv}}$ construction for Experiment 1 (n=6), with pairwise degrees $\ell \le 5$, equal to the $\mathcal{F}_{K,\mathrm{inv}}$ cutoff; the $\mathcal{F}_K$, $\mathcal{F}_{\mathrm{inv}}$, and $\mathcal{F}_0$ dictionaries (Table 3) therefore omit part of the targets' spectral support. All four estimators are applied to the same noisy observations of each target, with 30 repetitions per target.

- **Experiment 4** (adaptive): An $\mathcal{F}_{K,\mathrm{inv}}$ target drawn as in Experiment 3 (n=6, K^*=2, seed 42), with pairwise degrees $\ell \le 5$; its spectral support is contained in the pair dictionaries of both the K'=2 and K'=3 candidates. The adaptive estimator splits the data into two halves: the first half fits candidate estimators at each $K' \in \{1,2,3\}$ on the pure-mode dictionaries of Definition 38, the second half computes holdout risk for each candidate, the candidate with smallest holdout risk is selected, and the final estimator is refit on all N samples using the selected $\widehat{K}$ before test-set evaluation; the candidate curves and the empirical oracle in Figure 4 are likewise computed from fits on all N samples. Degree cutoffs are fixed a priori, $\ell_{\max,2}$=5 for pair modes and $\ell_{\max,3}$=1 for triplet modes, held constant across all N and applied per factor on the relative coordinates. The K'=3 candidate uses the full-support $(1,1)$ relative-coordinate block restricted to $\mathrm{range}(P_\Delta^\perp)$, a 72-dimensional pure subspace per triplet.

- **Experiment 5** (Maier–Saupe): a Maier–Saupe-type pair potential on $\mathrm{SO}(3)$,

$$f^*(\mathbf{g}) = -c_n \sum_{i<j} \big[P_2(x_{ij}) + 0.3\, P_4(x_{ij}) \big], \qquad x_{ij} = \cos\theta_{ij} = \tfrac{1}{2}\big(\mathrm{tr}(g_i^{-1} g_j) - 1\big),$$

where P_2 and P_4 are Legendre polynomials and θ_{ij} is the rotation angle of $g_i^{-1} g_j$ (not the angle between molecular directors of classical Maier–Saupe theory [23]). The constant $c_n > 0$ is determined from 5,000 independently sampled rotation tuples so that $\mathrm{Std}(f^*(X)) \approx 0.1$; the main experiment uses no three-body perturbation ($\lambda = 0$; see Appendix M). The target is globally invariant and pairwise (K=2), with per-pair Peter–Weyl support in degrees $\ell \le 4$.

Table 3: Experimental parameters across all five main experiments. Supplementary studies (Appendices L, M, and N) use the same template with parameter deviations stated inline.

Parameter	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
n	5	3–8	6	6	6, 10
K	2	2	2	K^*=2, $K_{\max}$=3	2
s	2	2	2	2	2
σ^2	0.1	0.1	0.1	0.1	0.1
N range	500–50K	500–50K	500–50K	500–50K	100–10K
N_{test}	5000	5000	5000	5000	5000
Repetitions	30	20	30	30	30
$\ell_{\max}$ ($\mathcal{F}_{K,\text{inv}}$)	5	5	5	5 (pairs), 1 (triplets)	5
$\ell_{\max}$ ($\mathcal{F}_K$)	3(1b)+2(2b)	3(1b)+2(2b)	1(1b)+1(2b)	—	1(1b)+1(2b)
$\ell_{\max}$ ($\mathcal{F}_{\text{inv}}$)	2	2	2	—	2
$\ell_{\max}$ ($\mathcal{F}_0$)	1	1	1	—	1
GT type	per-class	per-class	shared	shared	Maier–Saupe

Notation. For $\mathcal{F}_K$, entries "a(1b)+b(2b)" indicate cutoff a for one-body terms and cutoff b for two-body terms (since $\mathcal{F}_K$ has both). $\mathcal{F}_{K,\text{inv}}$ has no 1-body invariant terms, so a single cutoff suffices. For Experiment 4, the $\mathcal{F}_{K,\text{inv}}$ row lists the pair and triplet cutoffs $\ell_{\max,2}$, $\ell_{\max,3}$ of the adaptive candidates, which use the invariant-class pullback (Definition 38), so the $\mathcal{F}_K$ row is marked "—". For $\mathcal{F}_{\text{inv}}$ and $\mathcal{F}_0$, the listed $\ell_{\max}$ values are computational *safety caps* on the per-factor degree, not the active cutoff at each N: the binding threshold is the joint Casimir cutoff Λ_{cut}, set per experiment as described under "*Convention by estimator family*" in the Estimator implementations subsection.

K.3 Cross-experiment parameter table

K.4 Compute

All experiments were run on a single MacBook Pro with an Apple M1 Max chip and 64 GB of unified memory. The spectral-estimator experiments ran on CPU. The neural-bridge code (Appendices O–P) uses PyTorch with the Metal Performance Shaders (MPS) backend when available, and CPU otherwise. No external cluster was used. The spectral estimators are closed-form finite-dictionary estimators with the cutoff and shrinkage conventions described above; they require no iterative optimization. With the reported cutoffs and repetitions, each main experiment and appendix study—including the supplementary K=3 analysis (Appendix N), which uses higher body-order cutoffs and two-level regularization with a pure-S projection step—completes within approximately 12 hours of wall-clock time on this machine; the slower configurations are those that combine the largest sample size ($N = 50{,}000$), the highest body-order cutoff, and 20–30 repetitions.

L Smoothness Robustness

We check that the main experimental findings are not artifacts of the specific smoothness s=2.

L.1 Experiment 1 with $s = 3$

We repeat Experiment 1 with smoothness s=3 (all other parameters unchanged). At higher smoothness, the theory predicts steeper slopes: $|\alpha_{K,\text{inv}}| = 6/9 \approx 0.667$, $|\alpha_K| = 6/12 = 0.500$, $|\alpha_{\text{inv}}| = 6/18 \approx 0.333$, $|\alpha_0| = 6/21 \approx 0.286$. Figure 6 shows that the empirical slope ordering $|\hat{\alpha}_{K,\text{inv}}| > |\hat{\alpha}_K| > |\hat{\alpha}_{\text{inv}}| \gtrsim |\hat{\alpha}_0|$ is preserved: $|\hat{\alpha}_{K,\text{inv}}| \approx 0.63$, $|\hat{\alpha}_K| \approx 0.30$, $|\hat{\alpha}_{\text{inv}}| \approx 0.20$, and $|\hat{\alpha}_0| \approx 0.20$—with $\mathcal{F}_{\text{inv}}$ and $\mathcal{F}_0$ nearly tied, consistent with $\mathcal{F}_0$ being flagged pre-asymptotic in panel (b). The $\mathcal{F}_{K,\text{inv}}$ empirical slope is close to the theoretical prediction (within 0.04); for larger-D classes, the steeper theoretical slope at s=3 requires substantially larger N to observe, consistent with the wider pre-asymptotic regimes reported in Experiment 1.

L.2 Experiment 5 with $s = 3$, $n = 6$

We repeat the Maier–Saupe benchmark (Experiment 5) at s=3 with n=6. Figure 7a shows that the four-class hierarchy $\mathcal{F}_{K,\text{inv}} < \mathcal{F}_K < \mathcal{F}_{\text{inv}} < \mathcal{F}_0$ is preserved for $N \gtrsim 200$; at N=100 the

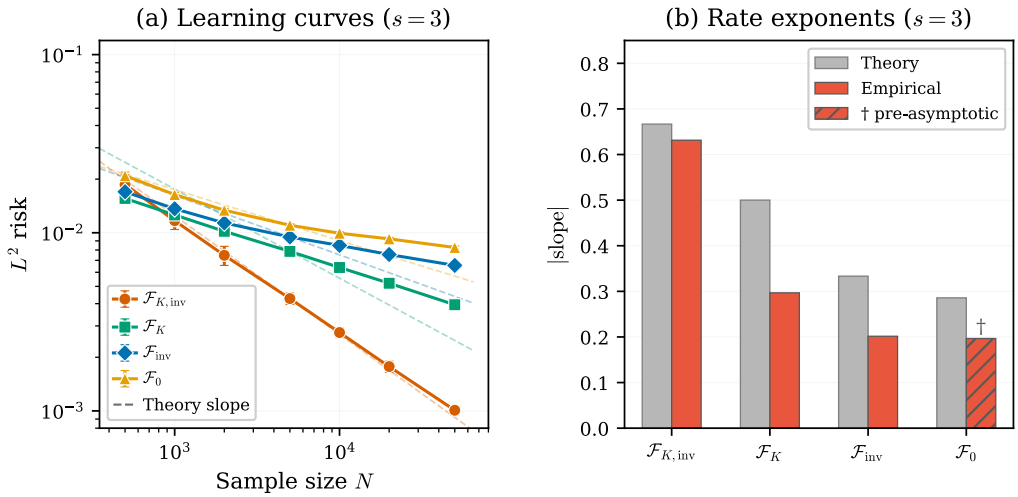

Figure 6: Supplementary smoothness robustness check for Experiment 1. At s=3, the empirical slope ordering is preserved. The $\mathcal{F}_{K,\mathrm{inv}}$ slope is within 0.04 of the theoretical prediction, while other classes show larger finite-sample deviations due to wider pre-asymptotic regimes at higher smoothness.

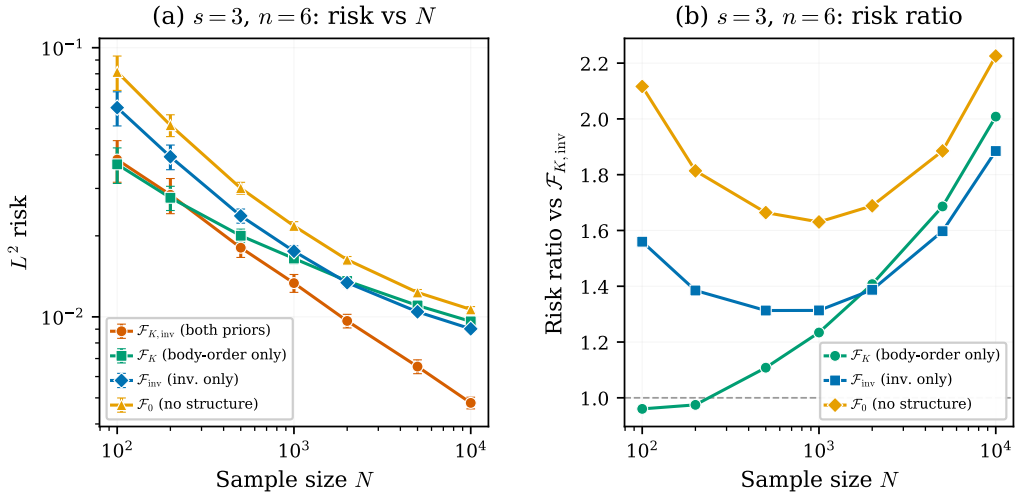

Figure 7: Supplementary smoothness robustness check for Experiment 5. At s=3, the four-class hierarchy is preserved. $\mathcal{F}_{K,\mathrm{inv}}$ and $\mathcal{F}_K$ become harder to distinguish at small N because higher smoothness makes all classes easier, compressing the risk ratios between adjacent classes.

$\mathcal{F}_K/\mathcal{F}_{K,\mathrm{inv}}$ ratio dips just below 1 (panel b), reflecting that higher smoothness s=3 makes the two lowest-risk classes nearly indistinguishable when data is very scarce—consistent with the rate formula, where a larger s increases $2s/(2s+D)$ for all classes but narrows the relative differences. The risk ratios (Figure 7b) are correspondingly smaller than at s=2 but maintain the same ordering from N=200 onward.

M Misspecification Robustness

We test whether the hierarchy is robust to mild violation of the pairwise assumption by adding a three-body perturbation with λ=0.3:

$$f^*_\lambda(\mathbf{g}) = c_{n,\lambda}\Big[-\sum_{i<j}\big(P_2(x_{ij}) + 0.3\,P_4(x_{ij})\big) + \lambda \sum_{i<j<k} P_2(x_{ij})\,P_2(x_{jk})\Big],$$

where x_{ij} is as in Appendix K and $c_{n,\lambda} > 0$ again normalizes the target standard deviation to approximately 0.1. Under this contamination, $f^*_\lambda \notin \mathcal{F}_{K,\mathrm{inv}}$ (it has a genuine three-body component),

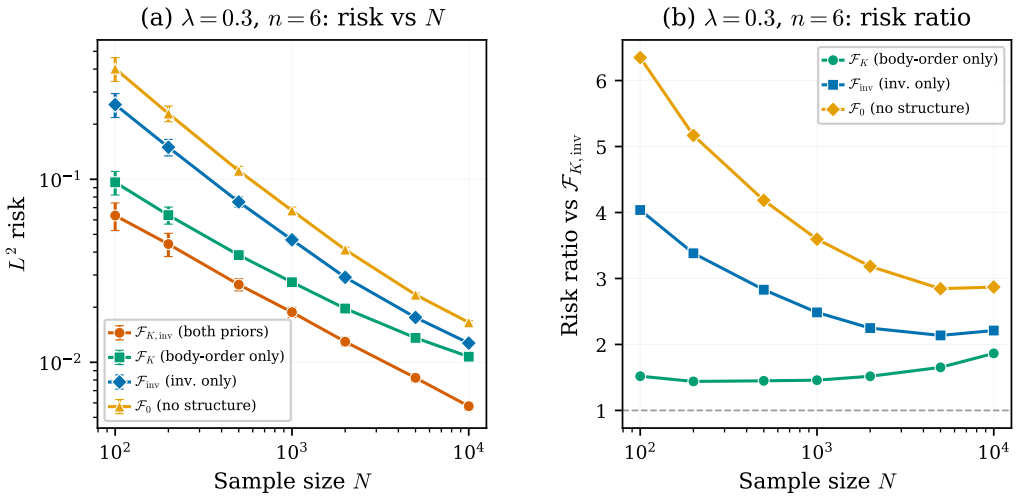

Figure 8: Supplementary misspecification robustness check for Experiment 5 at n=6. Under λ=0.3 three-body misspecification, the four-class hierarchy is preserved, with risk ratios close to the clean setting. The pairwise estimator $\mathcal{F}_{K,\mathrm{inv}}$ retains its advantage because the contamination is mild and the dominant signal remains pairwise.

so the pairwise estimator incurs a nonzero approximation bias. The question is whether this bias is small enough to preserve the hierarchy at practical sample sizes. Larger λ values would probe a different regime in which genuinely higher-order structure dominates the signal, a distinct question from hierarchy robustness.

M.1 Experiment 5 with $\lambda = 0.3$

Figure 8 shows the results at n=6: the four-class hierarchy $\mathcal{F}_{K,\mathrm{inv}} < \mathcal{F}_K < \mathcal{F}_{\mathrm{inv}} < \mathcal{F}_0$ is preserved, with risk ratios nearly identical to the clean (λ=0) setting. The pairwise estimator $\mathcal{F}_{K,\mathrm{inv}}$ retains its advantage because the contamination is mild and the dominant signal remains pairwise.

At n=10 (Figure 9), the hierarchy is again preserved; the gap between $\mathcal{F}_{K,\mathrm{inv}}$ and the growing-D classes $\mathcal{F}_{\mathrm{inv}}, \mathcal{F}_0$ widens from n=6 to n=10 (Table 4), while $\mathcal{F}_K$'s multiplier remains nearly n-independent—the n-dependent body-order benefit persists under misspecification.

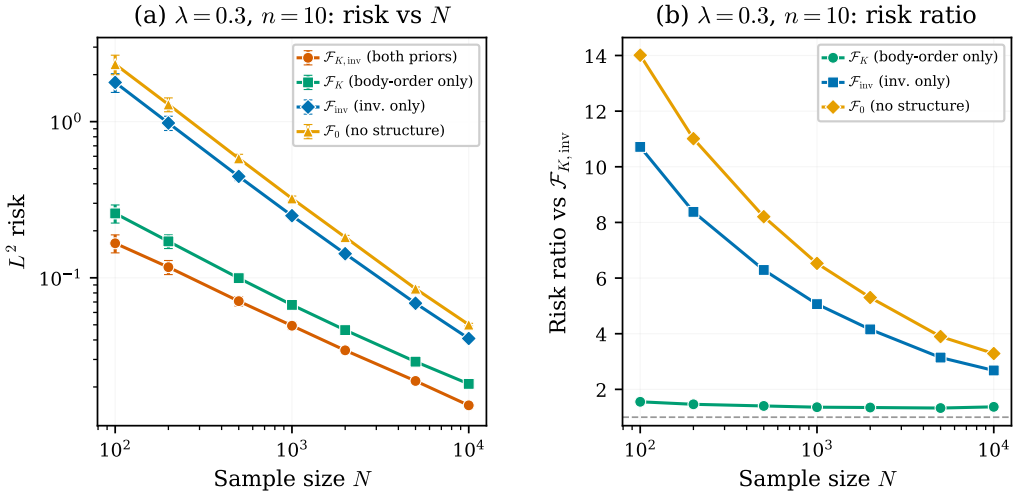

Figure 9: Supplementary misspecification robustness check for Experiment 5 at n=10. Under λ=0.3 three-body misspecification, the hierarchy remains preserved, with the gap between $\mathcal{F}_{K,\mathrm{inv}}$ and the growing-D classes ($\mathcal{F}_{\mathrm{inv}}, \mathcal{F}_0$) larger than at n=6.

Table 4: Supplementary robustness summary: sample-efficiency multipliers at anchor $N{=}500$ across five settings. Each entry shows how many times more samples a class needs to match $\mathcal{F}_{K,\mathrm{inv}}$ at $N{=}500$. Rows marked "(main)" are the Experiment 5 baselines reported in the main paper; other rows are the supplementary perturbations (this appendix). All multipliers shift by less than $0.3\times$ under $\lambda{=}0.3$ misspecification. At $s{=}3$, multipliers are smaller (higher smoothness compresses the sample-efficiency gaps) but ordering is preserved.

Setting		$\mathcal{F}_K$	$\mathcal{F}_{\mathrm{inv}}$	$\mathcal{F}_0$
$n{=}6$, $s{=}2$, $\lambda{=}0$	(main)	$2.2\times$	$4.8\times$	$8.3\times$
$n{=}6$, $s{=}2$, $\lambda{=}0.3$		$2.1\times$	$4.7\times$	$8.1\times$
$n{=}6$, $s{=}3$, $\lambda{=}0$		$1.4\times$	$1.9\times$	$3.1\times$
$n{=}10$, $s{=}2$, $\lambda{=}0$	(main)	$1.8\times$	$9.8\times$	$12.9\times$
$n{=}10$, $s{=}2$, $\lambda{=}0.3$		$1.8\times$	$9.6\times$	$12.7\times$

M.2 Robustness summary table

Table 4 summarizes the sample-efficiency multipliers at anchor $N{=}500$ across five settings.

The $\lambda{=}0.3$ rows are nearly identical to the clean rows in this setup, showing that this level of three-body contamination does not erase the practical advantage of the pairwise structural prior. The $s{=}3$ row shows smaller multipliers, consistent with the theory: higher smoothness makes all classes easier, compressing the sample-efficiency multipliers between classes. The ordering $\mathcal{F}_K < \mathcal{F}_{\mathrm{inv}} < \mathcal{F}_0$ (relative to $\mathcal{F}_{K,\mathrm{inv}}$) is preserved across all five settings.

N Supplementary $K{=}3$ Study: Pure-S Pullback at the First Nontrivial Order

The main-text experiments use $K{=}2$. Theorem 18 records that $P_\Delta^\perp$ acts as the identity on full-support inputs at $k{=}2$, so the pure-S projection becomes a *nontrivial* operator only at $k \geq 3$. This appendix reports a supplementary study at $K{=}3$ whose purpose is to check that the pure-S pullback construction is *operational* at the first nontrivial order, rather than to re-establish the asymptotic rate quantitatively (a task for which the main-text $K{=}2$ study is already well-suited).

We restrict the comparison to $\mathcal{F}_{K,\mathrm{inv}}$ vs $\mathcal{F}_K$ because this contrast isolates the effect of invariance and the pure-S pullback: both classes share the $K{=}3$ body-order truncation and differ only in whether diagonal invariance is imposed. Including $\mathcal{F}_{\mathrm{inv}}$ and $\mathcal{F}_0$ would test the full four-class hierarchy again, but it would not further isolate the nontrivial pure-S projection. We therefore keep this appendix as a targeted construction check rather than a second full hierarchy experiment; the main-text $K{=}2$ experiments already establish the $\mathcal{F}_{\mathrm{inv}}/\mathcal{F}_0$ ordering.

N.1 Setup

We set $n{=}5$, $K{=}3$, $s{=}5$, and $\sigma^2{=}0.1$. The smoothness $s{=}5$ places *both* $\mathcal{F}_{K,\mathrm{inv}}$ ($D_{\mathrm{eff}}{=}6$) and $\mathcal{F}_K$ ($D_{\mathrm{eff}}{=}9$) within the theorem's regime ($s > D_{\mathrm{eff}}/2$): the margins are 2 and 0.5, respectively.

Ground truth. As in Experiment 1, each estimator is evaluated on its own class-matched ground truth; this isolates per-class behavior while enabling a direct $\mathcal{F}_K/\mathcal{F}_{K,\mathrm{inv}}$ comparison of class-specific learning-curve decay. For $\mathcal{F}_{K,\mathrm{inv}}$, we generate random Sobolev-decay coefficients on each pair orbit ($\ell \leq 5$) and each triplet orbit ($\ell_1, \ell_2 \leq 2$); triplet coefficients are then passed through the pure-S projection $P_\Delta^\perp$ (Theorem 18), ensuring the triplet signal lies strictly in $\mathcal{V}_S$ for $|S|{=}3$ with no leakage into lower-order $\mathcal{V}_{S'}$ subspaces (pair-level coefficients are already in $\mathcal{V}_S$ for $|S|{=}2$ without projection, since $P_\Delta^\perp = \mathrm{id}$ at $k{=}2$). For $\mathcal{F}_K$, we use a standard ANOVA construction with body-order cutoffs $(\ell_{1\mathrm{b}}, \ell_{2\mathrm{b}}, \ell_{3\mathrm{b}}) = (2, 1, 1)$.

Estimator. Both estimators follow the spectral projection template of Appendix K, with level-matched Tikhonov regularization (each body-order level uses its own regularization strength, chosen to match that level's effective dimension). The $\mathcal{F}_{K,\mathrm{inv}}$ triplet estimator applies $P_\Delta^\perp$ to the raw coefficient estimates before reconstruction, preventing double-counting of 2-body content between

the pair and triplet levels. Estimator cutoffs are $(\ell_{\text{pair}}, \ell_{\text{trip}}) = (3, 1)$; intentionally lower than the ground-truth cutoffs so the estimator does not exhaust the signal in the tested N range.

Measurements. We sweep $N \in \{500, 1000, 2000, 5000, 10000, 20000, 50000\}$ and measure L^2 risk over $M{=}20$ independent replications; error bars in figures show ± 1 standard deviation, consistent with the main-text convention.

N.2 Results

Construction is operational. Figure 10(a) shows both estimators produce clean, monotonically decreasing risk curves across two orders of magnitude in N; neither exhibits plateau, instability, or non-monotone behavior. In particular, the $\mathcal{F}_{K,\text{inv}}$ curve does *not* flatten at large N—the concern that motivated this study in the first place—consistent with the intended role of the $P_\Delta^\perp$ projection in removing lower-order leakage between pair and triplet levels.

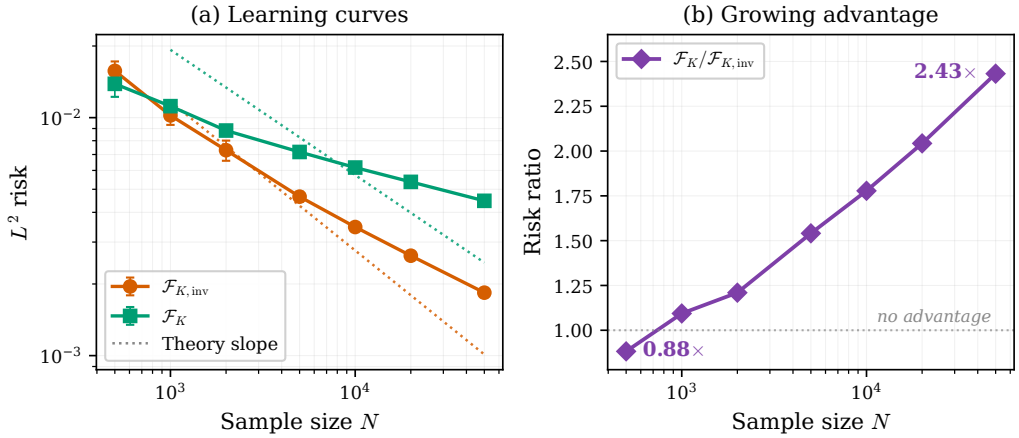

Figure 10: Supplementary $K{=}3$ study ($n{=}5$, $s{=}5$, $\sigma^2{=}0.1$; 20 repetitions, error bars show ± 1 standard deviation, matching the main-text convention). **Panel (a):** Learning curves for $\mathcal{F}_{K,\text{inv}}$ (pure-S pullback, orange) and $\mathcal{F}_K$ (green). Short dotted segments indicate the theoretical asymptotic slopes, anchored at $N{=}5000$ for visual reference only. **Panel (b):** Risk ratio $R(\mathcal{F}_K)/R(\mathcal{F}_{K,\text{inv}})$ vs. N, growing monotonically from $0.88\times$ at $N{=}500$ to $2.43\times$ at $N{=}50{,}000$. At $N{=}500$ the ratio is slightly below unity; this is a finite-sample effect discussed in the text and does not contradict the theorem.

Pure-S pullback yields a growing advantage. The risk ratio $R(\mathcal{F}_K)/R(\mathcal{F}_{K,\text{inv}})$ grows monotonically from $1.09\times$ at $N{=}1000$ to $2.43\times$ at $N{=}50{,}000$ (Figure 10(b)). This monotone growth is consistent with the 3-unit effective-dimension reduction afforded by invariance (from $D_{\mathcal{F}_K} = 3K$ to $D_{\mathcal{F}_{K,\text{inv}}} = 3(K{-}1)$; Corollary 2), which predicts $\mathcal{F}_{K,\text{inv}}$ to decay strictly faster than $\mathcal{F}_K$ asymptotically (exponent-magnitude gap $|\alpha_{K,\text{inv}}| - |\alpha_K| = 0.099$). The observed growth is larger than the exponent gap alone would predict over two orders of magnitude in N, reflecting additional finite-sample differences between the two estimator families; we do not attempt to decompose these effects here. The crossover at $N \approx 10^3$ suggests that, in this setup, the invariance advantage becomes visible within the tested sample range, rather than only in the far-asymptotic regime.

Small-N behavior. At $N{=}500$ the two estimators have comparable risk, with $\mathcal{F}_K$ marginally lower (1.39×10^{-2} vs. 1.57×10^{-2}; ratio 0.88). One possible explanation involves the class-matched target structure: the two target generators place energy differently across pair and triplet components, and at very small N these finite-dimensional constants and regularization effects can dominate the asymptotic dimension advantage. This effect is finite-sample in nature and does not contradict Theorem 1, which is an asymptotic statement; the $\mathcal{F}_K$ advantage disappears by $N{=}1000$ and inverts thereafter.

On slope fitting. A direct log-log slope fit to the $\mathcal{F}_{K,\text{inv}}$ data gives $|\hat{\alpha}| \approx 0.46$, below the theoretical magnitude $|\alpha_{K,\text{inv}}| = 0.625$. We do not interpret these observed slopes as quantitative validation

of the asymptotic rate: as stated at the opening of this appendix, the purpose of this study is to check that the pure-S pullback construction is operational at the first nontrivial order ($k \geq 3$), not to re-establish the asymptotic exponent quantitatively. The informative observables are the qualitative signatures—monotone decrease, growing $\mathcal{F}_K/\mathcal{F}_{K,\mathrm{inv}}$ ratio, no plateau—consistent with the theorem's predictions.

N.3 Summary

The $K{=}3$ supplementary study provides three qualitative observations: (i) the pure-S pullback construction (Theorem 18) is numerically stable and operational in the first regime where it is nontrivial ($P_\Delta^\perp \neq \mathrm{id}$); (ii) the risk advantage of $\mathcal{F}_{K,\mathrm{inv}}$ over $\mathcal{F}_K$ grows monotonically with N, with the crossover at $N \approx 10^3$ suggesting that, in this setup, the invariance advantage becomes visible within the tested sample range; and (iii) we interpret the apparent slope deviations from theory as finite-sample behavior rather than as quantitative evidence about the asymptotic exponent. We do not claim quantitative rate validation at $K{=}3$ in the tested sample range; strict asymptotic slope matching would require substantially larger N and is a natural direction for future empirical work.

O Neural Bridge on the Maier–Saupe Task

Purpose and scope. This appendix provides a small, architecture-facing bridge from our spectral-projection experiments to learned models. The goal is neither to validate Theorem 1 at the neural level nor to study modern equivariant neural networks. Instead, we check whether the four-class structural hierarchy predicted by the theory—and corroborated at the spectral-estimator level in Section 4—remains visible in simple prior-aware learners trained end-to-end on the same Maier–Saupe benchmark. Throughout, we refer to our four neural learners as $\mathcal{F}_0$-*like*, $\mathcal{F}_{\mathrm{inv}}$-*like*, $\mathcal{F}_K$-*like*, and $\mathcal{F}_{K,\mathrm{inv}}$-*like*, paralleling the four structural classes in the main text.

Models. All four models are built from comparable MLP primitives and are designed to differ primarily in which of the two structural priors they encode. $\mathcal{F}_0$-like is an unconstrained MLP taking the concatenation of the n absolute rotations as input, thus encoding neither body-order truncation nor global invariance. $\mathcal{F}_{\mathrm{inv}}$-like takes relative rotations R_{1j} (for $j{=}2,\ldots,n$) as input, encoding global-frame invariance but not body-order truncation. $\mathcal{F}_K$-like is a pair aggregator that sums a shared pair network over the unordered pairs (R_i, R_j); it encodes body-order truncation ($K{=}2$) but not invariance. $\mathcal{F}_{K,\mathrm{inv}}$-like combines both priors: a shared pair network is applied to the relative rotations $R_{ij} = R_i^\top R_j$ and aggregated by summation over unordered pairs. The four models are parameter-matched by tuning the pair-network hidden width so that their total parameter counts agree within 8–9%, reducing the possibility that differences in risk are driven merely by capacity rather than by structural inductive bias.

Protocol. We use the exact Maier–Saupe target family from Experiment 5 (Section 4), evaluated at $n{=}6$ and $n{=}10$. For each sample size $N \in \{500, 1000, 2000, 5000, 10000, 20000\}$ and each of 10 random seeds, we generate three independent rotation samples: a training set of size N, a validation set of size 2000, and a test set of size 5000. Training and validation use noisy labels drawn from the supervised observation model, so that the validation signal used for early stopping has the same noise structure as training; final test error is measured *once* against the underlying clean Maier–Saupe target, with the test set never used for model selection. Optimization uses Adam, 500 epochs, early stopping with patience 50 on the validation loss, gradient clipping, and a single learning-rate schedule shared across models. No per-model hyperparameter tuning is performed. Because the same optimization protocol is used for all four learners, the bridge compares structural priors under a common training budget rather than maximizing each architecture's absolute performance.

What the bridge reports. We report the mean clean-target test MSE across the 10 seeds at the validation-selected checkpoint, together with its empirical standard deviation. We also report log-log slopes $\hat{\alpha}$ obtained from a least-squares fit of $\log \mathrm{MSE}$ against $\log N$ for each model, as a compact summary of how steeply each model's risk decays. We emphasize that these slopes are descriptive summaries of the bridge curves and are *not* interpreted as empirical rate exponents; the interpretation is qualitative throughout.

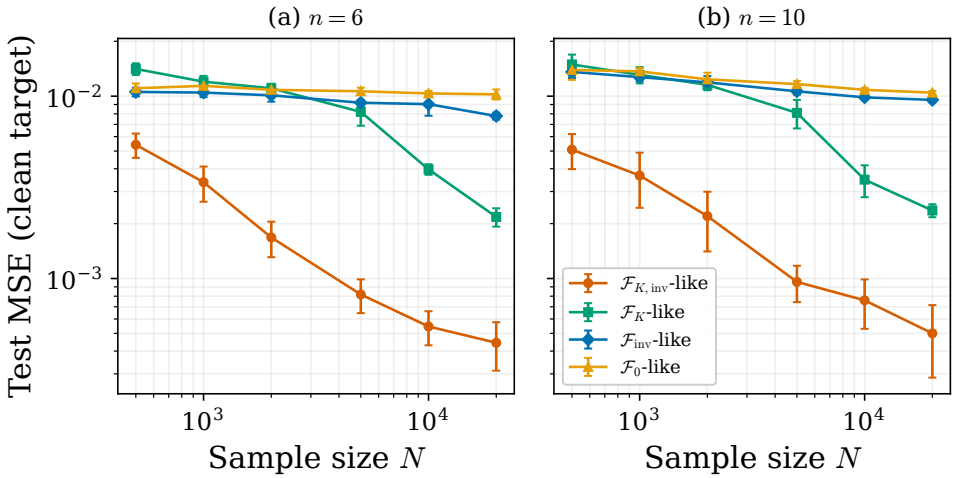

Figure 11: **Neural bridge on the Maier–Saupe task at n=6 (left) and n=10 (right).** Four parameter-matched learners encode none, one, or both of the two structural priors; they parallel the four structural classes of the main theory. Training uses a train/val/test protocol; test MSE is evaluated once against the underlying clean target, with the test set never used for model selection. Error bars show ± 1 standard deviation across 10 random seeds. The $\mathcal{F}_{K,\mathrm{inv}}$-like learner achieves the lowest risk at every sample size; the strict four-class ordering emerges at moderate-to-large N.

Table 5: **Summary of the neural bridge.** Log-log slopes $\hat{\alpha}_c$ for each neural learner $c \in \{\mathcal{F}_0, \mathcal{F}_{\mathrm{inv}}, \mathcal{F}_K, \mathcal{F}_{K,\mathrm{inv}}\}$-like are least-squares fits across $N \in \{500, \ldots, 20000\}$ on the mean curves over 10 seeds. "$\mathcal{F}_{K,\mathrm{inv}}$ best" reports the number of sample sizes (out of six) at which $\mathcal{F}_{K,\mathrm{inv}}$-like is the minimum-risk model. "Strict ordering" is the smallest N from which the full ordering $\mathcal{F}_{K,\mathrm{inv}}$-like $<$ $\mathcal{F}_K$-like $<$ $\mathcal{F}_{\mathrm{inv}}$-like $<$ $\mathcal{F}_0$-like holds uniformly on the means. The final column reports the endpoint clean-test-risk ratio between $\mathcal{F}_0$-like and $\mathcal{F}_{K,\mathrm{inv}}$-like at N=20,000. Slopes and endpoint ratios are descriptive summaries of the neural bridge experiment, not neural minimax-rate estimates.

n	$\hat{\alpha}_0$	$\hat{\alpha}_{\mathrm{inv}}$	$\hat{\alpha}_K$	$\hat{\alpha}_{K,\mathrm{inv}}$	$\mathcal{F}_{K,\mathrm{inv}}$ best	Strict ordering	$\mathcal{F}_0/\mathcal{F}_{K,\mathrm{inv}}$ at N=20,000
6	-0.03	-0.08	-0.49	-0.71	6/6	from N=5000	23×
10	-0.08	-0.10	-0.52	-0.66	6/6	from N=2000	21×

Results. Figure 11 and Table 5 summarize the bridge. Three qualitative features mirror the structural predictions of the main theory. First, the $\mathcal{F}_{K,\mathrm{inv}}$-like learner is the lowest-risk model at every tested sample size, at both n=6 and n=10 (Table 5). The endpoint gap is substantial: at N=20000, the clean-target test MSE of $\mathcal{F}_{K,\mathrm{inv}}$-like is 23× smaller than that of $\mathcal{F}_0$-like at n=6 and 21× smaller at n=10. Second, the body-order axis is clearly visible. The $\mathcal{F}_K$-like learner, which encodes body-order truncation alone, decays much more steeply than the two non-body-order classes ($\hat{\alpha} \approx -0.5$, versus $|\hat{\alpha}| \lesssim 0.1$ for $\mathcal{F}_0$-like and $\mathcal{F}_{\mathrm{inv}}$-like). This descriptive slope contrast mirrors the qualitative signature of the theory: removing high body-order levels reduces the rate-governing dimension by $3(n-K)$, which should appear as a steeper learning curve when the model can exploit the prior. Third, the strict four-class ordering $\mathcal{F}_{K,\mathrm{inv}}$-like $<$ $\mathcal{F}_K$-like $<$ $\mathcal{F}_{\mathrm{inv}}$-like $<$ $\mathcal{F}_0$-like emerges at moderate-to-large N (from N=5000 at n=6, from N=2000 at n=10). The ordering fails at the smallest sample sizes only because $\mathcal{F}_K$-like temporarily sits above $\mathcal{F}_{\mathrm{inv}}$-like (and, at the very smallest N, above $\mathcal{F}_0$-like as well): the $\mathcal{F}_K$-like pair network takes higher-dimensional (6D) pair inputs and must learn a single shared pair function applied across all $\binom{n}{2}$ pairs, which requires a minimum amount of data before its body-order structural advantage is realized. Once the shared pair representation is learned, $\mathcal{F}_K$-like's steeper learning curve overtakes $\mathcal{F}_{\mathrm{inv}}$-like and the ordering stabilizes.

Limitations and what we do *not* claim. The bridge is deliberately scoped. We do *not* claim that these neural learners attain the minimax rates of Theorem 1; the observed log-log slopes are descriptive summaries, not empirical rate estimates, and we make no direct comparison between

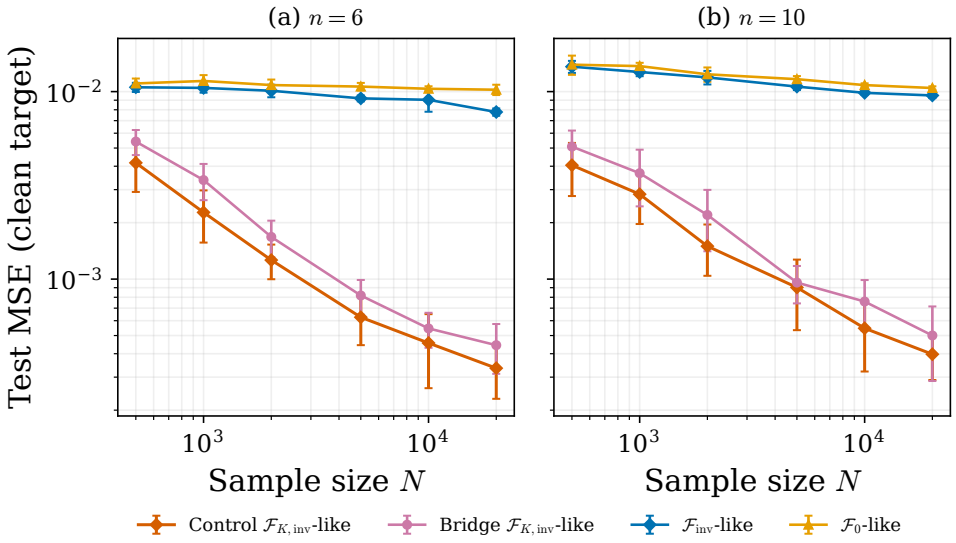

Figure 12: **Architectural control for the $\mathcal{F}_{K,\mathrm{inv}}$-like neural bridge at n=6 (left) and n=10 (right).** The control variant (absolute rotation input, internal R_{ij} construction, pair-swap symmetrization, sum-pool over unordered pairs) is parameter-matched to, and trained under the identical train/val/test protocol as, the $\mathcal{F}_{K,\mathrm{inv}}$-like bridge learner of Appendix O. At every tested sample size and both system sizes, the mean control curve lies below the mean bridge $\mathcal{F}_{K,\mathrm{inv}}$-like curve. $\mathcal{F}_0$-like and $\mathcal{F}_{\mathrm{inv}}$-like are shown for context. Error bars: ± 1 standard deviation across 10 random seeds.

$\hat{\alpha}$ and the theoretical exponents. We do *not* claim that this is a study of modern equivariant neural architectures: the four models are simple MLP-based learners that encode the priors in minimal form, and the result should be read as a qualitative architecture-facing consistency check rather than a benchmark of state-of-the-art equivariant networks such as MACE [3] or NequIP [4]. We also do *not* claim any quantitative gap-widening behavior as n grows; the bridge reproduces the qualitative four-class hierarchy at both n=6 and n=10, but its protocol is not designed to resolve rate-level n-dependence precisely. A systematic study of whether expressive equivariant neural architectures provably attain the minimax rates established in Theorem 1 remains an open question for the equivariant learning community, in line with the discussion in Section 5.

P Architectural Control for the $\mathcal{F}_{K,\mathrm{inv}}$-Like Neural Bridge

Purpose and scope. In the bridge of Appendix O, global invariance is implemented by preprocessing inputs into relative rotations before they enter the network, a convention chosen for simplicity and parity across the four models. This appendix provides a small architectural control that reimplements $\mathcal{F}_{K,\mathrm{inv}}$-like so that the relevant symmetries are enforced inside the architecture rather than by external preprocessing before the network. The purpose is narrow: to check that the $\mathcal{F}_{K,\mathrm{inv}}$-like bridge result does not depend on the specific feature-preprocessing choice of Appendix O. As in the bridge, this is not a study of modern equivariant neural networks.

Design. The control variant takes absolute rotations $R_1, \ldots, R_n$ as input. For every unordered pair $\{i, j\}$, relative rotations $R_{ij} = R_i^\top R_j$ and $R_{ji} = R_j^\top R_i$ are computed inside the forward pass. A shared pair network is applied to both R_{ij} and R_{ji}, their outputs are averaged to enforce pair-swap symmetry by construction, and the resulting per-pair features are aggregated by summation over unordered pairs, yielding a scalar output. By construction the control variant is globally invariant, pair-structured, and unordered-pair symmetric. Its parameter count is matched to the $\mathcal{F}_{K,\mathrm{inv}}$-like bridge learner of Appendix O. The training protocol (train/val/test split, 10 seeds, 500 epochs, patience 50, clean-target test evaluation) is identical.

Results. Figure 12 summarizes the control. Under the same train/val/test protocol as Appendix O, the control variant remains comparable to, and modestly improves upon, the $\mathcal{F}_{K,\mathrm{inv}}$-like bridge learner at both tested system sizes. The mean control curve lies below the mean bridge $\mathcal{F}_{K,\mathrm{inv}}$-like curve at every tested sample size ($12/12$ points across $n \in \{6, 10\}$), with typical improvements around 20% and a minimum of approximately 6% at $n{=}10, N{=}5000$. The two curves also have very similar descriptive log-log slopes (control $\hat{\alpha} = -0.69$ versus bridge -0.71 at $n{=}6$; control -0.65 versus bridge -0.66 at $n{=}10$), consistent with a roughly parallel downward shift rather than a qualitatively different decay pattern. We read this as evidence against the $\mathcal{F}_{K,\mathrm{inv}}$-like bridge result being solely an artifact of the specific feature-preprocessing choice in Appendix O: the same qualitative curve and comparable magnitude persist when global invariance and pair-swap symmetry are enforced inside the architecture rather than via external preprocessing.

Limitations and what we do *not* claim. The control uses a single architectural variant and is intended only to check robustness of the $\mathcal{F}_{K,\mathrm{inv}}$-like bridge result to one specific feature-preprocessing choice. It is not a substitute for a full architecture search or hyperparameter comparison; it is a targeted sanity check. We do not interpret it as evidence about modern equivariant neural architectures, and we do not claim quantitative rate attainment.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction identify the four Sobolev function classes corresponding to the combinations of global invariance and body-order truncation, their matching minimax rates under class-specific smoothness assumptions (Theorem 1), the additive reduction in effective dimension (Corollary 2), and fixed-family adaptive selection (Proposition 3). Section 4 reports finite-sample comparisons, including sample-efficiency multipliers on a Maier–Saupe-type benchmark, while Section 5 distinguishes theoretical guarantees from empirical observations.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 states the estimator, smoothness/body-order, empirical breadth, and domain limitations. The matching upper bounds are established for spectral projection; KRR achievability and guarantees for practical equivariant networks remain open. At $s{=}2$ only $\mathcal{F}_{K,\mathrm{inv}}$ satisfies the theorem's smoothness condition, and the $K{=}3$ study is a construction check rather than rate validation. Section 4 and Appendix K also explain that finite dictionaries and their approximation errors can affect empirical risk comparisons. Extensions to $\mathrm{SO}(m)^n$ are presented as future work.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Theorem 1 states the class-specific conditions $s > D_{\text{eff}}/2$ and the minimax rates; Proposition 3 gives the fixed-family oracle inequality and a conditional true-order candidate risk statement, without asserting minimax-rate adaptation. Proofs are provided in Appendices F, G, H, I, and J. The relative-coordinate pullback, pure-S projection, mode counting, and Sobolev norm comparisons are developed in Appendices B–D.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Appendix K provides all spectral-estimator definitions, the synthetic ground-truth generators, the Maier–Saupe-type target, numerical protocols, and Table 3 with the sample sizes, noise, cutoff, test-size, and repetition settings. Appendices L, M, N, and O–P document supplementary and neural studies. Public scripts, archived results, and the plotting entry point at `https://github.com/zheshuoli/invariance-bodyorder-minimax` let readers regenerate the reported figures and tables from the supplied data, or rerun the experiments under the documented configurations.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All experiment scripts, generated result JSON files, figures, and plotting code are publicly available at `https://github.com/zheshuoli/invariance-bodyorder-minimax` under the MIT license. The repository README documents dependencies, the recorded computing environment, random seeds, quick checks, full-run commands, and how to regenerate the paper figures and tables with `python plots/make_figures.py`. Experiments use synthetic rotations and analytically defined targets, so no external dataset download is required; Appendix K documents the protocols and parameter settings.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

Justification: Appendix K specifies the Haar design, ground-truth generators, noise variance, sample sizes, test-set size, repetition counts, per-class Fourier dictionaries, Casimir/degree cutoffs, and experiment-specific Sobolev-weighted shrinkage. Section 4.4 uses a 50/50 fit–holdout split followed by full-data refitting. Appendices O–P describe the neural train/validation/test splits, ten seeds, Adam training, 500 epochs, patience-50 early stopping, and the architectural control. The public code and README give executable configurations.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The five main spectral experiments use 20 or 30 repetitions per configuration (Table 3); Experiment 3 averages over three targets with 30 repetitions per target. Learning-curve error bars indicate ± 1 empirical standard deviation across repetitions, as defined in Appendix K and the figure captions. The neural bridge and architectural control report ± 1 standard deviation over ten seeds (Appendices O–P). Derived slope, risk-ratio, and selection-frequency panels report descriptive point estimates; formal confidence intervals or significance tests are not claimed.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Appendix K (Compute) reports a single MacBook Pro with an Apple M1 Max chip and 64 GB unified memory, without an external cluster. Spectral-estimator experiments run on CPU with finite dictionaries and closed-form coefficient updates; the neural-bridge scripts use PyTorch and select the Metal Performance Shaders (MPS) backend when available, otherwise CPU, and train with Adam. The public README at

`https://github.com/zheshuoli/invariance-bodyorder-minimax` gives approximate full-run times by experiment and instructions for regenerating figures directly from archived results.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This is a study of nonparametric estimation on $\mathrm{SO}(3)^n$ using synthetic data and a Maier–Saupe-type rotational pair potential. It does not involve human participants, personal data, or sensitive datasets. The research is consistent with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Section 5 (Broader significance) discusses potential scientific benefits for understanding sample efficiency in symmetry-aware and body-ordered learning. The contribution is foundational statistical theory, with no deployed decision-making system or personal data, and no direct negative societal impact is identified. Any indirect benefits or risks would depend on downstream applications, which are outside the scope of this work.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The public repository contains research code, synthetic experimental results, and figures; it does not release pretrained high-risk models, scraped personal data, or other assets requiring misuse safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: No external datasets or pretrained models are used. The implementation depends on NumPy and SciPy (BSD 3-Clause licenses), Matplotlib (Matplotlib license, based on the PSF license), and PyTorch (BSD-style license); dependency and environment information is given in the public README and `requirements.txt` at `https://github.com/zheshuoli/invariance-bodyorder-minimax`. The Maier–Saupe inspiration is credited to its original physics reference [23]. The newly released experiment code is licensed separately under MIT.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The authors release the experiment scripts, 17 archived result JSON files, figure-generation script, and generated figures at `https://github.com/zheshuoli/invariance-bodyorder-minimax`. The repository contains a README with installation and reproduction commands, dependency information, and a MIT LICENSE file. No data involving human subjects are included.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve human subjects, so IRB approval is not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: No LLM is used as an important, original, or non-standard component of the core method. The mathematical analysis, finite spectral estimators, and reported neural bridge do not use an LLM.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.